

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:56:27 ; Search time 4995 Seconds
(without alignments)
11483.810 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 171
Sequence: 1 atgaaaaagtgataaccgc.....catggttgatcaacatctc 171

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pi:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pin:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1971	100.0	1971	6	AX006271	AX006271 Sequence
2	1971	100.0	217420	1	BSUB0017	Z99120 Bacillus su
3	807.8	41.0	1958	6	AX431809	AX431809 Sequence
4	238.2	12.1	265118	1	CNSPAX06	AJ248288 Pyrococcus
5	238.2	12.1	265118	6	AX041922	AX041922 Sequence
6	227.2	11.5	1896	6	AX116209	AX116209 Sequence
7	227.2	11.5	1896	6	E16634	E16634 Pyrococcus
8	227.2	11.5	233000	1	AP000003	AP000003 Pyrococcus
9	218.2	11.1	12882	1	AE013176	AE013176 Thermoana
10	192.2	9.8	477	6	AX436451	AX436451 Sequence
11	191.2	9.7	10256	1	AE001879	AE001879 Deinococc
12	161.8	8.2	251700	1	AP000062	AP000062 Aeropyrum
13	159.6	8.1	9531	1	AE010617	AE010617 Fusobacte
14	158.2	8.0	11233	1	AE009798	AJ248287 Pyrococcus
15	158	8.0	304290	1	CNSPAX05	AX041921 Sequence
16	158	8.0	349980	6	AX041921	AX041921 Pyrococcus
17	150.6	7.6	257000	1	AP000002	AP000002 Pyrococcus
18	149.8	7.6	1869	6	AR095647	AR095647 Sequence
19	149.8	7.6	1869	6	AR136115	AR136115 Sequence
20	143.4	7.3	268050	1	AP000983	AP000983 Sulfolobu
21	143.4	7.3	12358	1	AE006863	AE006863 Sulfolobu
22	143	7.3	12358	1	AP000994	AP000994 Thermopla
23	132.6	6.7	182440	1	AE005114	AE005114 Halobacte
24	130.4	6.6	9825	1	AP000064	AP000064 Aeropyrum
25	130.4	6.6	247695	1	AE010156	AE010156 Pyrococcus
26	128.4	6.5	13354	1	AE013111	AE013111 Thermoana
27	128	6.5	11370	1	AE013111	AX436423 Sequence
28	82.4	4.2	585	6	AX436423	AX436423 Sequence
29	75.2	3.8	139203	2	AC088311	AC088311 Homo sapi
30	73	3.7	12098	1	AE012209	AE012209 Xanthomon
31	65	3.3	1998	3	AF059278	AF059278 Entamoeba
32	64.8	3.3	2400	1	BSTHRC	AF059278 Entamoeba
33	64.2	3.3	12621	1	AE008621	AE008621 Rickettsi
34	62.8	3.2	1992	3	AF059279	AF059279 Entamoeba
35	59.2	3.0	2642	1	AF508804	AF508804 Xanthomon
36	57	2.9	11388	1	AE011750	AE011750 Xanthomon
37	54.4	2.8	7218	6	I66494	I66494 Sequence 14
38	48.4	2.5	11505	1	AE012120	AE012120 Xanthomon
39	46.4	2.4	1797	6	AX431923	AX431923 Sequence
40	46.4	2.4	12345	1	AE006757	AE006757 Sulfolobu
41	46.2	2.3	316900	1	TACID3	AL445065 Thermopla
42	46	2.3	2325	8	AF082514	AF082514 Trichophy
43	45.4	2.3	13928	1	AF004038	AF004038 Xylella f
44	45.2	2.3	782	11	PM11H10G	AL684260 Penicilli
45	45.2	2.3	10393	1	AE011651	AE011651 Xanthomon

ALIGNMENTS

RESULT 1
AX006271
LOCUS AX006271
DEFINITION Sequence 1 from Patent WO9903984.
ACCESSION AX006271
VERSION AX006271.1 GI:9929124
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL

AX006271
Sequence 1 from Patent WO9903984.
AX006271.1 GI:9929124
Bacillus subtilis.
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1 (bases 1 to 1971)
Estell,D.A.
Proteases from gram-positive organisms
Patent: WO 9903984-A 1 28-JAN-1999;
GENENCOR INTERNATIONAL B V (NL); ESTELL DAVID A (US)

1971 bp
DNA
linear
PAT 24-AUG-2000

FEATURES

Location/Qualifiers

source

1. .1971
/organism="Bacillus subtilis"
/db_xref="taxon:1423"

BASE COUNT 558 a 431 c 524 g 458 t

ORIGIN

Query Match 100.0%; Score 1971; DB 6; Length 1971;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGAAAGCTGATACACCGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 60
|||||
Db 1 ATGAAAGCTGATACACCGACGACATCACAGCGATTGTCTCTGTGACCGATCCTCAA 60
|||||

QY 61 TAGGCCCCAGAGCGGTACCGTCCGCATATGTAATACAAATCAAGTAAATCAAGAGAAGAT 120
|||||
Db 61 TAGGCCCCAGAGCGGTACCGTCCGCATATGTAATACAAATCAAGTAAATCAAGAGAAGAT 120
|||||

QY 121 TCGTATACATCAATATATGATCTATGAACGAAACGGGAGGATCTGCTTCTTGGACA 180
|||||
Db 121 TCGTATACATCAATATATGATCTATGAACGAAACGGGAGGATCTGCTTCTTGGACA 180
|||||

QY 181 CATGGAGAAAGCGAAGCACCACCAAGATGGTCTCCGGAGCGGCGACGCTTGCCTTT 240
|||||
Db 181 CATGGAGAAAGCGAAGCACCACCAAGATGGTCTCCGGAGCGGCGACGCTTGCCTTT 240
|||||

QY 241 ATTTCTGATGAGAGCGGATCGCGCACAGCTTTATATCATGAGCACTGAAGCGGAGAA 300
|||||
Db 241 ATTTCTGATGAGAGCGGATCGCGCACAGCTTTATATCATGAGCACTGAAGCGGAGAA 300
|||||

QY 301 GCAAGAAACTGACTGATATCCATATGCGGTGTCGAGCGGCTATGTCCTCCGACGGT 360
|||||
Db 301 GCAAGAAACTGACTGATATCCATATGCGGTGTCGAGCGGCTATGTCCTCCGACGGT 360
|||||

QY 361 GAATCGATTCTGGTCACTATCAGTTTGGGAGAGGGGAAAGCACTGATGACCGGAGAAAA 420
|||||
Db 361 GAATCGATTCTGGTCACTATCAGTTTGGGAGAGGGGAAAGCACTGATGACCGGAGAAAA 420
|||||

QY 421 ACAGAGGACGACGATGAACCTTGTGAAGTCAAGGCCCTCTCTCAAAACGGGAGGC 480
|||||
Db 421 ACAGAGGACGACGATGAACCTTGTGAAGTCAAGGCCCTCTCTCAAAACGGGAGGC 480
|||||

QY 481 AAAGGGCTGACGAGAGGTGGGTATGCCAGCTTGTCTGTGTCAGCGTAAAGTCCGGTGA 540
|||||
Db 481 AAAGGGCTGACGAGAGGTGGGTATGCCAGCTTGTCTGTGTCAGCGTAAAGTCCGGTGA 540
|||||

QY 541 ATGAAAGCTGACAACTGATCAAGCTGATCATGGTGTATCCTTTCTCTGACGGC 600
|||||
Db 541 ATGAAAGCTGACAACTGATCAAGCTGATCATGGTGTATCCTTTCTCTGACGGC 600
|||||

QY 601 AAATGGCTTGTCTCAGCTAATTTAACTGAACAGATGATGCCAGCAAGCGCGATGAT 660
|||||
Db 601 AAATGGCTTGTCTCAGCTAATTTAACTGAACAGATGATGCCAGCAAGCGCGATGAT 660
|||||

QY 661 GTTTACATATGTCAGTGGAGTCTGAGATCTTAAAGCAGGTTACACCTCATCGGGCTCA 720
|||||
Db 661 GTTTACATATGTCAGTGGAGTCTGAGATCTTAAAGCAGGTTACACCTCATCGGGCTCA 720
|||||

QY 721 TTCGGATCAAGCTCATTTTACCAGACGGAGGATCTGCTTTGCTTGGAAATGAAAAG 780
|||||
Db 721 TTCGGATCAAGCTCATTTTACCAGACGGAGGATCTGCTTTGCTTGGAAATGAAAAG 780
|||||

QY 781 GAATATAAGATGCTACGCTCTCAAGGGGTGGCTCTATGATATCAACAAAGCGCCCTC 840
|||||
Db 781 GAATATAAGATGCTACGCTCTCAAGGGGTGGCTCTATGATATCAACAAAGCGCCCTC 840
|||||

QY 841 ACATGCTTACTGAGATGCTGACGCTTCAATTTAGCGGATGCGGTGATGAGATTCATTG 900
|||||
Db 841 ACATGCTTACTGAGATGCTGACGCTTCAATTTAGCGGATGCGGTGATGAGATTCATTG 900
|||||

QY 901 ATCGGTGGTGTGAACAGCGCCGATTTGGACAAAGACAGCAAGGGTTTATGTCTATC 960
|||||

RESULT 2
BSUB0017/c

901 ATCGGTGGTGTGAACAGCGCCCGATTGTCACAAAGGACAGCCAAAGGTTTATGTCTATC 960
|||||
QY 961 GGCACAGATCAAGCGCAGTACGGGCATCTATTATATTTTCGATTTGAAGCCCTTGTGATCCG 1020
|||||
Db 961 GGCACAGATCAAGCGCAGTACGGGCATCTATTATATTTTCGATTTGAAGCCCTTGTGATCCG 1020
|||||

QY 1021 ATTCTGTCTGAAAAAGAGTACATCAATAGCTTTTCTTTTCCACTGATGACAGCACTTT 1080
|||||
Db 1021 ATTCTGTCTGAAAAAGAGTACATCAATAGCTTTTCTTTTCCACTGATGACAGCACTTT 1080
|||||

QY 1081 ATTGCCAGTGTGACAAAGCGGACAGACCGAGTGAAGCTTTTACAGTATCCCGTTGGACAG 1140
|||||
Db 1081 ATTGCCAGTGTGACAAAGCGGACAGACCGAGTGAAGCTTTTACAGTATCCCGTTGGACAG 1140
|||||

QY 1141 GAAGAGAAAACAGCTGACTGGCGGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATA 1200
|||||
Db 1141 GAAGAGAAAACAGCTGACTGGCGGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATA 1200
|||||

QY 1201 CCTGAAGAGATTCAATATGCTACAGAAAGCGGCTGATGGTGAACGGCTGGCTGATGAGG 1260
|||||
Db 1201 CCTGAAGAGATTCAATATGCTACAGAAAGCGGCTGATGGTGAACGGCTGGCTGATGAGG 1260
|||||

QY 1261 CCTGCACAAATGAAAGGTGAGACAAATATCCACTTATTTCTTAAACATACACGGCGTCCG 1320
|||||
Db 1261 CCTGCACAAATGAAAGGTGAGACAAATATCCACTTATTTCTTAAACATACACGGCGTCCG 1320
|||||

QY 1321 CATATGATGTACGGACATACATATTTTCATGAGTTTTCAGGTGCTGGCGGCGAAAGGATAC 1380
|||||
Db 1321 CATATGATGTACGGACATACATATTTTCATGAGTTTTCAGGTGCTGGCGGCGAAAGGATAC 1380
|||||

QY 1381 CGGTCGCTTATATCAATCCGAGAGAAAGCCAGCTACGGGCGAGAAATTTGTGAATGCG 1440
|||||
Db 1381 CGGTCGCTTATATCAATCCGAGAGAAAGCCAGCTACGGGCGAGAAATTTGTGAATGCG 1440
|||||

QY 1441 GTCAGAGAGATTATGGGGAAAGGATTATGACGATGTGATGACGGCTGTGGATGAGCT 1500
|||||
Db 1441 GTCAGAGAGATTATGGGGAAAGGATTATGACGATGTGATGACGGCTGTGGATGAGCT 1500
|||||

QY 1501 ATCAAGAGATCCGCTATTTGATCTTAAGCGGCTGGTGTACACGGGCGGAGCTTACCGA 1560
|||||
Db 1501 ATCAAGAGATCCGCTATTTGATCTTAAGCGGCTGGTGTACACGGGCGGAGCTTACCGA 1560
|||||

QY 1561 GGTTTTATGACCAACTGGATCGTGGGCGAGACAAACCGCTTTAAAGCTGCGGTTACCCAG 1620
|||||
Db 1561 GGTTTTATGACCAACTGGATCGTGGGCGAGACAAACCGCTTTAAAGCTGCGGTTACCCAG 1620
|||||

QY 1621 CGCTCGATATCAAAATTTGGATCAGCTTTTACGGCGTCACTGATATCGGCTATTTCTTACA 1680
|||||
Db 1621 CGCTCGATATCAAAATTTGGATCAGCTTTTACGGCGTCACTGATATCGGCTATTTCTTACA 1680
|||||

QY 1681 GACTGGCAGCTTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGCTCTCCT 1740
|||||
Db 1681 GACTGGCAGCTTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGCTCTCCT 1740
|||||

QY 1741 TTAATATCCAGCAACGTTGGAGACACCGCTTTTGTACTGATGATGCGGAGCGGATGAC 1800
|||||
Db 1741 TTAATATCCAGCAACGTTGGAGACACCGCTTTTGTACTGATGATGCGGAGCGGATGAC 1800
|||||

QY 1801 CGATGCCGATCGAGAGCGGAGAGCTGCTTTTATCGCTCTGAAAAAATTTGGCAAGAA 1860
|||||
Db 1801 CGATGCCGATCGAGAGCGGAGAGCTGCTTTTATCGCTCTGAAAAAATTTGGCAAGAA 1860
|||||

QY 1861 ACCAAGCTGTCCGTTTCCGAATGATCGCAATTTATACGACCGGACACCCCAAGA 1920
|||||
Db 1861 ACCAAGCTGTCCGTTTCCGAATGATCGCAATTTATACGACCGGACACCCCAAGA 1920
|||||

QY 1921 CAGCGGATCAAGCGCTGAATATATACGCTCATGCTTGTGATCAACATCTC 1971
|||||
Db 1921 CAGCGGATCAAGCGCTGAATATATACGCTCATGCTTGTGATCAACATCTC 1971
|||||

LOCUS	BSUB0017	217420 bp	DNA	linear	BCT 26-NOV-1997
DEFINITION	Bacillus subtilis complete genome (section 17 of 21): from 3197001 to 3414420.				
ACCESSION	Z99120 AL009126				
VERSION	Z99120.1 GI:2635613				
KEYWORDS					
SOURCE	Bacillus subtilis.				
ORGANISM	Bacillus subtilis.				
REFERENCE	Bacteria: Firmicutes; Bacillales; Bacillaceae; Bacillus.				
AUTHORS	Kunst, F., Ogasawara, N., Moszer, I., Albertini, A. M., Alloni, G., Azevedo, V., Bertsch, M. G., Bessieres, P., Bolotin, A., Borchert, S., Boriss, R., Boursier, L., Brans, A., Braun, M., Brignell, S. C., Bron, S., Brouillet, S., Brusch, C. V., Caldwell, B., Capuano, V., Carter, N. M., Choi, S. K., Codani, J. J., Connerton, I. F., Cummings, N. J., Daniel, R. A., Denizot, F., Devine, K. M., Dusterhoft, A., Ehrlich, S. D., Emerson, P. T., Entian, K. D., Errington, J., Fabret, C., Ferrari, E., Foulger, D., Fritz, C., Fujita, M., Fujita, Y., Fuma, S., Galizzi, A., Galleron, N., Ghim, S. Y., Glaser, P., Goffeau, A., Gollightly, E. J., Grandi, G., Guisepi, G., Guy, B. J., Haga, K., Halech, J., Harwood, C. R., Henaut, A., Hilbert, H., Holsappel, S., Hosono, S., Hullo, M. F., Ito, Y., Jones, J., Joris, B., Karamata, D., Kasahara, Y., Kjaer-Blanchard, M., Klein, C., Kobayashi, Y., Koetter, P., Koningsstein, G., Krogh, S., Kumano, M., Kurita, K., Lapidus, A., Lardinois, S., Lauber, J., Lazarevic, V., Lee, S. M., Levine, A., Liu, H., Masuda, S., Maue, C., Medigue, C., Medina, N., Mellado, R. P., Mizuno, M., Moestl, D., Nakai, S., Noback, M., Noone, D., O'Reilly, M., Ogawa, K., Ogiwara, A., Oudega, B., Park, S. H., Parro, V., Pohl, T. M., Portetle, B., Porwollik, S., Prescott, A. M., Presecan, E., Pujic, P., Purnelle, B., Rapoport, G., Rey, M., Reynolds, S., Rieger, M., Rivolta, C., Rocha, E., Roche, B., Rose, M., Sadale, Y., Sato, T., Scanlan, E., Schleich, S., Schroeter, R., Scoffone, F., Sekiguchi, J., Sekowska, A., Seror, S. J., Serror, P., Shin, B. S., Soldo, B., Sorokin, A., Tacconi, E., Takagi, T., Takahashi, H., Takemaru, K., Takeuchi, M., Tanakoshi, A., Tanaka, T., Terpstra, P., Tognoni, A., Tosato, V., Uchiyama, S., Vandenbol, M., Vannier, F., Vassarotti, A., Viari, A., Wambutt, R., Wendenbol, M., Vannier, F., Vassarotti, A., Winters, P., Wipat, A., Yamamoto, H., Yamane, K., Yasumoto, K., Yata, K., Yoshida, K., Yoshikawa, H. F., Zumstein, E., Yoshikawa, H. and Danchin, A.				
TITLE	The complete genome sequence of the gram-positive bacterium				
JOURNAL	Bacillus subtilis				
MEDLINE	Nature 390 (6657), 249-256 (1997)				
PUBMED	98044033				
REFERENCE	9384377				
AUTHORS	2 (bases 1 to 217420)				
TITLE	Kunst, F., Ogasawara, N., Yoshikawa, H. and Danchin, A.				
JOURNAL	Direct Submission				
FEATURES	Submitted (18-NOV-1997) I. Moszer, A. Danchin, Institut Pasteur, Regulation de l'Expression Genetique, 28 rue du Docteur Roux, 75724 Paris Cedex 15, FRANCE. E-mail: moszer@pasteur.fr, adanchin@pasteur.fr Phone: +33 (0)1 45 68 84 41, Fax: +33 (0)1 45 68 89 48				
source	Location/Qualifiers				
	1. .217420				
	/organism="Bacillus subtilis"				
	/strain="168"				
	/db_xref="taxon:1423"				
	complement(3. .1277)				
	/gene="yulB"				
	complement(3. .1277)				
	/gene="yulB"				
	/function="unknown"				
	/note="similar to L-rhamnose isomerase"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAB15107.1"				
	/db_xref="GI:2635614"				
	/db_xref="SPTREMBL:O05264"				
	/translation="MTIKANDSAKQAYEKWIDVEEALROLEQVPSIHWCQDDIE GFENKLSGSDIVTGNYPGAQTPPELRDLKALSLIPGKRVNLHAIYETNRE AVERDLKPOHFENKWKWAKNLGLDFNPTFLSHEKAADGLTSHDPDPIREFWIRH CIACRRIGEYFGKLGTPCLTNIPDGYKDIPDRTPRRLKESLDRIFSEISEQ				
gene	HNLDSIESKLGIGSESYYVGSHEFYLAYALTNHKLCLLDTGTHFHTPTVTSNKISSML LYTDKLAHVSRRPVWSDHVVLDDELRETALEIVRNHAKLEKVAIGLDFFDASINRY AAWTIGTRNMKIALLYALLPGLNGLKQEQEGRTERIALMEEFKTYPPFGAIWDSYCE OMGVPVKEAWLYDIKEYBQQVLLKRAKSSPIV"				
	complement(1303. .1617)				
	/gene="yulD"				
	complement(1303. .1617)				
	/gene="yulD"				
	/function="unknown"				
	/note="similar to hypothetical proteins"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAB15108.1"				
	/db_xref="GI:2635615"				
	/db_xref="SPTREMBL:O05263"				
	/translation="MKRKASIMFVHODKYEEYKQRHDDIWPEMAALKAHGAHHSIF LDFTGRLFAYLEIEDEKRWKMDTEVCYRWKWSMAPLMKNTSDFSPVAILKEVYF LD"				
	complement(1635. .3092)				
	/gene="yulC"				
	complement(1635. .3092)				
	/gene="yulC"				
	/function="unknown"				
	/note="similar to rhamnulokinase"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAB15109.1"				
	/db_xref="GI:2635616"				
	/db_xref="SPTREMBL:O05262"				
	/translation="MIYTAIDVGASSGRIMVGELENGKLDIOETHRFANGFSQDGHG LWDIDHLKILQLOKQVKTGTYGHCFTVIGTWDVYVLLDEKGRLEAISYDRRT DHTDKLEHTLSKAAIYOKTQIOFQFNTIYOLFEDRELLKTKDKIMMIDYLGICYL TGKAVETITVNTQLLAVSTGNLDPELLEAVSLEQOQFAPLTPGCEGLKRNWEP DYDLPACKVMTVATHDFTASAVIARPGVNDGWAYISSTWSLIGVKNKTPILTDLLEN NYTTERGANNTRIRLKNITIGHMWIOEYVQKQLOADISFOQLAEEAKKTPFQOQFINLD KRFNLPENMIKEIOHQYCTOKIPTAGELACCIYSNLAIYIAIAIKELETTIKPPI EGFHIGGARNDNFQNLQTDMSKAVYAGPIEATATGNLLMQMTAAKEVKDIKEARQ VVRNSFPKVFPTPKDIDRSTIIQSFQQTVLKALS"				
	complement(3097. .3873)				
	/gene="yulB"				
	complement(3097. .3873)				
	/gene="yulB"				
	/function="unknown"				
	/note="similar to transcriptional regulator (Deor family)"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAB15110.1"				
	/db_xref="GI:2635617"				
	/db_xref="SWISS-PROT:O05261"				
	/translation="MLVAERQOKIIVEIVNMRSSIRVSELSDFSVTEETIRDLKLE KEHLRSRGGAVSIOQKSEIHFSEIREITNVIEKKAIAHEAAKYVKSQDRILDAST TAWYMAKILPDIELTVITNSKAAITELSKENISVISTGILLESKLSFAGPLAERSL ETYHVNKTFLSCKGPDVHNGSDSNEMOALLKRMERSDOTILMADSSKWKGNREFSH TASQDVSRRLTDSGLDPSAKALEDKKVKVTAVPLSKRG"				
	complement(3930. .5999)				
	/gene="yuxG"				
	complement(3930. .5999)				
	/gene="yuxG"				
	/function="unknown"				
	/note="alternate gene name: yulA similar to sorbitol-6-phosphate 2-dehydrogenase"				
	/codon_start=1				
	/transl_table=11				
	/protein_id="CAB15111.1"				
	/db_xref="GI:2635618"				
	/db_xref="SWISS-PROT:P40747"				
	/translation="MVKHIWDSERAAQLPKGVVEELVYRNLIGSDRTVCNWGGNTSM KTKEDFRGETIEVNMWVKGSDSLATMKAHNPSGLKDDIRPLIKRQDQPMDEMYDYL SHCMDSKHPSPSIETLLHAFYKHYVDHTHPDAIISICCADNQIAEDYIGNRFVW VPYVPGETLSKMTAEFVANNPHALVLMKHLVTVGTSETCYQKTIISIOEAEQY INDRINQHEVGGKRYQPLPEKRRQKILAGINPVIRGAVSEKKMILSYDDHDDVLEF VNSVOAPALSQIGAAACPDHLVTRKRPVLYIDNPETQDVHKLADLIKSGVETFTSEIQ AYFTRNQDQDGQIFESAPRVLIPGIMVNTYQSKVAMSQVSGALYRAIYAVMKGATAL				

		QGFVSLHENESVHYWPELYKLTLLAPPEAFESFRKVALITGGAGGIGSAACRRFAAE GGHVIALDNLIEGAKIAGENIDAYKGRAMAVKMDVTKEEDVQSFARAAIAYGGID IIVNAGLATISPFETSLKWNLNMLVLTGYFLVAREAFQMKHQNRGSGVMVFYGS KNSYVAGKNASYGSVKALETHLARCIAAEGGEGFGRVNSVLPDAVLOGSAIWGSSWR ERAAAYGIEPDQLEEYHVKRKTALLVNIYPEDIAEAIAFFASSKAEKTTGCMITVDVG VPAAPTR"
gene		complement(6109. .8125) /gene="tipb" complement(6109. .6124) /gene="tipb" complement(6137. .8125) /gene="tipb" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15112.1" /db_xref="GI:2635619" /db_xref="SWISS-PROT:P39217"
terminator		/translation="MGKTIQWIKPSISKPLIAEFLAVLILPVGVLAYFSYOSAMNAL DRELIISAKGNVEELNLOKLEKDKVAIDYISQVTDKDKILLGNKTLILKFKFOYT TLNDVGAITYAASEEDKLYKIPDGPVKGFDPTRGWYQVAEAGQVAFSPSPYTD TGDIWVITSKLQKDGSGVIALDNLDELVTASKRIKGEKGFATITGNKKYIAHPTI KPGTIGSDWNTNOVYSKGESEFVTEFEGEKKMAETTNKLTGWKTAGTTFVSELQDAS SPVLNTAVILICSVIIGGILLYIIRAITKPKRLVSTSAKISGDLUTEVIDHSKN EFGQGESFNEMSASVIGVIOITSVENVASSELTAASAAOTSKATSHITLAEQF SDNGEASEKLETSNHLISQNMEGISKVAQASSTITKSIQSEAAAGSGEKLEHTVG QMKITDQSVQKAEAVGLKTSQDITISNTINGTADQTNELLANAAIEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLQEIIVREITSLSMFQSVNHYEKGELQITDQ TAESFKQIYENTQISGELONLNATVEQLSAGQSVSSVEDISAVAKESSAGIQDIA ASAEQLASMEIISSTLANMAEELQDITKKFKIES" complement(8204. .10224) /gene="mcpA" complement(8204. .8228) /gene="mcpA" complement(8239. .10224) /gene="mcpA"
gene		/function="chemotaxis deficiency toward glucose and alpha-methyl-glucoside" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15113.1" /db_xref="GI:2635620" /db_xref="SWISS-PROT:P39214"
terminator		/translation="MKTIQLIKQRSITRKLIVSFLSILIIIPVILAIAPAYOSASSL DROMMGSALENVQOLNEINTSIGEKENSADYFSEWLTKKYNAKNSASIREKFSQYI SINKDVESIYDQKHGHPYRDLPMPSGYNPVRDWYKKAANKGVITDPYKTAS TNTWVITYADQTKGSGVIAINMTIENLLTKTKVNIQTGYAFITMKDKKVAHPNE QSGTELGDWIDKMLSDAKDGFQYTMGDKKRMKAFDTNKLTKGKIGTGMVLDIEIHEAA OPVLHIALVILAAIIIGIIVYITLIRSIPTPKQLGVSSKRISGCDITETIDIRSKD ELGELGKSFNNASSLSLIIHAIOISDVNVAASSELTAASAAOTSKATSHITLAEQF SNGNEKQENIETAAEHYIYQMKDGLTNMAQASEVITDSSVQSTETASGEGKLVHQTG QMNVIDKSVKEAQVPLEGTFSKSDITNLRVINGIADQTNELLANIAEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLIIEIVKEINTSLGMPFSVQNEVQTGLDITDK
gene		complement(6109. .8125) /gene="tipb" complement(6109. .6124) /gene="tipb" complement(6137. .8125) /gene="tipb" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15112.1" /db_xref="GI:2635619" /db_xref="SWISS-PROT:P39217"
terminator		/translation="MGKTIQWIKPSISKPLIAEFLAVLILPVGVLAYFSYOSAMNAL DRELIISAKGNVEELNLOKLEKDKVAIDYISQVTDKDKILLGNKTLILKFKFOYT TLNDVGAITYAASEEDKLYKIPDGPVKGFDPTRGWYQVAEAGQVAFSPSPYTD TGDIWVITSKLQKDGSGVIALDNLDELVTASKRIKGEKGFATITGNKKYIAHPTI KPGTIGSDWNTNOVYSKGESEFVTEFEGEKKMAETTNKLTGWKTAGTTFVSELQDAS SPVLNTAVILICSVIIGGILLYIIRAITKPKRLVSTSAKISGDLUTEVIDHSKN EFGQGESFNEMSASVIGVIOITSVENVASSELTAASAAOTSKATSHITLAEQF SDNGEASEKLETSNHLISQNMEGISKVAQASSTITKSIQSEAAAGSGEKLEHTVG QMKITDQSVQKAEAVGLKTSQDITISNTINGTADQTNELLANAAIEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLQEIIVREITSLSMFQSVNHYEKGELQITDQ TAESFKQIYENTQISGELONLNATVEQLSAGQSVSSVEDISAVAKESSAGIQDIA ASAEQLASMEIISSTLANMAEELQDITKKFKIES" complement(8204. .10224) /gene="mcpA" complement(8204. .8228) /gene="mcpA" complement(8239. .10224) /gene="mcpA"
gene		/function="chemotaxis deficiency toward glucose and alpha-methyl-glucoside" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15113.1" /db_xref="GI:2635620" /db_xref="SWISS-PROT:P39214"
terminator		/translation="MKTIQLIKQRSITRKLIVSFLSILIIIPVILAIAPAYOSASSL DROMMGSALENVQOLNEINTSIGEKENSADYFSEWLTKKYNAKNSASIREKFSQYI SINKDVESIYDQKHGHPYRDLPMPSGYNPVRDWYKKAANKGVITDPYKTAS TNTWVITYADQTKGSGVIAINMTIENLLTKTKVNIQTGYAFITMKDKKVAHPNE QSGTELGDWIDKMLSDAKDGFQYTMGDKKRMKAFDTNKLTKGKIGTGMVLDIEIHEAA OPVLHIALVILAAIIIGIIVYITLIRSIPTPKQLGVSSKRISGCDITETIDIRSKD ELGELGKSFNNASSLSLIIHAIOISDVNVAASSELTAASAAOTSKATSHITLAEQF SNGNEKQENIETAAEHYIYQMKDGLTNMAQASEVITDSSVQSTETASGEGKLVHQTG QMNVIDKSVKEAQVPLEGTFSKSDITNLRVINGIADQTNELLANIAEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLIIEIVKEINTSLGMPFSVQNEVQTGLDITDK
gene		complement(6109. .8125) /gene="tipb" complement(6109. .6124) /gene="tipb" complement(6137. .8125) /gene="tipb" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15112.1" /db_xref="GI:2635619" /db_xref="SWISS-PROT:P39217"
terminator		/translation="MGKTIQWIKPSISKPLIAEFLAVLILPVGVLAYFSYOSAMNAL DRELIISAKGNVEELNLOKLEKDKVAIDYISQVTDKDKILLGNKTLILKFKFOYT TLNDVGAITYAASEEDKLYKIPDGPVKGFDPTRGWYQVAEAGQVAFSPSPYTD TGDIWVITSKLQKDGSGVIALDNLDELVTASKRIKGEKGFATITGNKKYIAHPTI KPGTIGSDWNTNOVYSKGESEFVTEFEGEKKMAETTNKLTGWKTAGTTFVSELQDAS SPVLNTAVILICSVIIGGILLYIIRAITKPKRLVSTSAKISGDLUTEVIDHSKN EFGQGESFNEMSASVIGVIOITSVENVASSELTAASAAOTSKATSHITLAEQF SDNGEASEKLETSNHLISQNMEGISKVAQASSTITKSIQSEAAAGSGEKLEHTVG QMKITDQSVQKAEAVGLKTSQDITISNTINGTADQTNELLANAAIEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLQEIIVREITSLSMFQSVNHYEKGELQITDQ TAESFKQIYENTQISGELONLNATVEQLSAGQSVSSVEDISAVAKESSAGIQDIA ASAEQLASMEIISSTLANMAEELQDITKKFKIES" complement(8204. .10224) /gene="mcpA" complement(8204. .8228) /gene="mcpA" complement(8239. .10224) /gene="mcpA"
gene		/function="chemotaxis deficiency toward glucose and alpha-methyl-glucoside" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15113.1" /db_xref="GI:2635620" /db_xref="SWISS-PROT:P39214"
terminator		/translation="MKTIQLIKQRSITRKLIVSFLSILIIIPVILAIAPAYOSASSL DROMMGSALENVQOLNEINTSIGEKENSADYFSEWLTKKYNAKNSASIREKFSQYI SINKDVESIYDQKHGHPYRDLPMPSGYNPVRDWYKKAANKGVITDPYKTAS TNTWVITYADQTKGSGVIAINMTIENLLTKTKVNIQTGYAFITMKDKKVAHPNE QSGTELGDWIDKMLSDAKDGFQYTMGDKKRMKAFDTNKLTKGKIGTGMVLDIEIHEAA OPVLHIALVILAAIIIGIIVYITLIRSIPTPKQLGVSSKRISGCDITETIDIRSKD ELGELGKSFNNASSLSLIIHAIOISDVNVAASSELTAASAAOTSKATSHITLAEQF SNGNEKQENIETAAEHYIYQMKDGLTNMAQASEVITDSSVQSTETASGEGKLVHQTG QMNVIDKSVKEAQVPLEGTFSKSDITNLRVINGIADQTNELLANIAEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLIIEIVKEINTSLGMPFSVQNEVQTGLDITDK
gene		complement(6109. .8125) /gene="tipb" complement(6109. .6124) /gene="tipb" complement(6137. .8125) /gene="tipb" /codon_start=1 /transl_table=11 /product="methyl-accepting chemotaxis protein" /protein_id="CAB15112.1" /db_xref="GI:2635619" /db_xref="SWISS-PROT:P39217"
terminator		/translation="MGKTIQWIKPSISKPLIAEFLAVLILPVGVLAYFSYOSAMNAL DRELIISAKGNVEELNLOKLEKDKVAIDYISQVTDKDKILLGNKTLILKFKFOYT TLNDVGAITYAASEEDKLYKIPDGPVKGFDPTRGWYQVAEAGQVAFSPSPYTD TGDIWVITSKLQKDGSGVIALDNLDELVTASKRIKGEKGFATITGNKKYIAHPTI KPGTIGSDWNTNOVYSKGESEFVTEFEGEKKMAETTNKLTGWKTAGTTFVSELQDAS SPVLNTAVILICSVIIGGILLYIIRAITKPKRLVSTSAKISGDLUTEVIDHSKN EFGQGESFNEMSASVIGVIOITSVENVASSELTAASAAOTSKATSHITLAEQF SDNGEASEKLETSNHLISQNMEGISKVAQASSTITKSIQSEAAAGSGEKLEHTVG QMKITDQSVQKAEAVGLKTSQDITISNTINGTADQTNELLANAAIEAARAGEYG RGFSVAEVRKLAQVSADSAKEIEGLQEIIVREITSLSMFQSVNHYEKGELQITDQ TAESFKQIYENTQISGELONLNATVEQLSAGQSVSSVEDISAVAKESSAGIQDIA ASAEQLASMEIISSTLANMAEELQDITKKFK

QY	1321	CATATGATGATCGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGCGGAAAGATAC	1380	Db	1	ATGAGACAGCTGATACCGGAAAGACCTCATCAAGCTTGTATCCATTACCGATCCCCAG	60
Db	113474	CATATGATGATCGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGCGGAAAGATAC	113415	QY	61	TACGCCCCAGACGGTACCGCTCCCGCATATGTAAATACAGTAATCAAGAGAAGAT	120
QY	1381	GGGTGCTTTATATCAATCCGAGAGGAGCCAGCTACGGCCAGGAATTTGTGAATGG	1440	Db	61	TATTCGCCGATGGTGCAAAATTCGCTACGCTCAAAACAAAGTAACGAAAACAGGAC	120
Db	113414	GGGTGCTTTATATCAATCCGAGAGGAGCCAGCTACGGCCAGGAATTTGTGAATGG	113355	QY	121	TCGTATACATCAATATATGATCTATGAACGAAACGGGAGGATCTGTTCCTTGGACA	180
QY	1441	GTACAGGAGATTATGGGGAAAGATTATGACGATGTGATCAGGCTGTGATGAGGCT	1500	Db	121	TCATATGATTCGCATATCATGATCTACGACCGCGGAAACAAAGCTTCGGTGCATGGACG	180
Db	113354	GTACAGGAGATTATGGGGAAAGATTATGACGATGTGATCAGGCTGTGATGAGGCT	113295	QY	181	CATGGAGAAAGGAAAGACCGACCCCAAGATGTCTCCGAGCGGGCGACGCTTGCTCTT	240
QY	1501	ATCAACAGAGATCCGCATATTCATCCTAAGCGGCTCGGTGTCAACGGGGGAAAGCTACGGA	1560	Db	181	TTTGGAAAGGAGAAACAGCATCCCGCTGCTCCGCGCAAAATACCTTGTGATTT	240
Db	113294	ATCAACAGAGATCCGCATATTCATCCTAAGCGGCTCGGTGTCAACGGGGGAAAGCTACGGA	113235	QY	241	ATTTCGATCGAAGAGCGCATGGCGCACAGCTTTATATCATGAGCACTGAAGCGGAGAA	300
QY	1561	GGTTTATGACCAACTGGATCTCGGGAGAGCAACCGCTTTAAAGCTCCCGTTACCCAG	1620	Db	241	AGTCCAAATCGGCA---GGAAACCGCACAAATTTAGTGATAGCGCGCGGGGGAGAA	297
Db	113234	GGTTTATGACCAACTGGATCTCGGGAGAGCAACCGCTTTAAAGCTCCCGTTACCCAG	113175	QY	301	GCAAGAAAATGACTGATATCCCATATGGCGTCAAGCGCTATGGTCCCGGAGCGT	360
QY	1621	CGCTCGATATCAAAATGGATCAGCTTTACGGCGTCAAGTATGATCGGCTATTTCTTTACA	1680	Db	298	GCAGAAAAGTGACCGATATTCCTATGATGATACAGCTGAAATGGTCCCGGACGCG	357
Db	113174	CGCTCGATATCAAAATGGATCAGCTTTACGGCGTCAAGTATGATCGGCTATTTCTTTACA	113115	QY	361	GAATCGATTCTGGTCACTATPAGTTTGGGAGAGGGGAAAGCAATTGATGACCGGAAAAA	420
QY	1681	GACTGGCAGCTTGACGATCACATGTTTGAGGACACAGAAAGCTCTGGACCGGCTCCT	1740	Db	358	AGTCAATTCGTTTTCAGTGAAGCTGACAAAAAGAGAAAGCGTCGACGATGAGAAGAA	417
Db	113114	GACTGGCAGCTTGACGATCACATGTTTGAGGACACAGAAAGCTCTGGACCGGCTCCT	113055	QY	421	ACAGACGAGACAGCTATGAACCTGTGAAGTGCAGGCGCTCTCTCAACAGGGACGCG	480
QY	1741	TTAAATACGACGAAACCTGGAGACACCGCTTTTGATCTGATCGGAGCGGGATGAC	1800	Db	418	ACCGGATTGAAGACGACCAACCGCTGGAAGTGCATCTTTGTCATATAAGACGACGCG	477
Db	113054	TTAAATACGACGAAACCTGGAGACACCGCTTTTGATCTGATCGGAGCGGGATGAC	112995	QY	481	AAAGGGTGACGAGAGGTGCTATGCGCAGCTTGTGCTGTGTCAGCGTAAAGTCGGGTGAG	540
QY	1801	CGATGCCGATCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGGAA	1860	Db	478	CAAGGTTTAAAGGGGAAAGTATACGACGCTTGTCTCTCAAGTGAACACCGGGGAA	537
Db	112994	CGATGCCGATCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGGAA	112935	QY	541	ATGAAAGAGCTGACAGCTACAAAGCTGATGATGCTGCTGCTCTCTCTCTCTCTCTCT	600
QY	1861	ACCAAGCTTGCTGCTTTCCGATGATCGCAAAATTTATCAGCAGCGGACACCCCAAGA	1920	Db	538	ATGAAACAGCTGACCGATCTTGAACGCGACCATTTACGCCATGATCTTACCTGCGGGA	597
Db	112934	ACCAAGCTTGCTGCTTTCCGATGATCGCAAAATTTATCAGCAGCGGACACCCCAAGA	112875	QY	601	AAATGGCTTCTTTCTCAGCTAAATTTAACTGAAACAGATGATGCCAGCAAGCCGATGAT	660
QY	1921	CAGCGGATCAAGCGCTGAATATATCAGCTCATGTTTGTGATCAACATCTC	1971	Db	598	GATCAGATTGCATTTTCGCGCAATCAACGGATATGCGTGT-----GAACGAT	645
Db	112874	CAGCGGATCAAGCGCTGAATATATCAGCTCATGTTTGTGATCAACATCTC	112824	QY	661	GTTCATATAATGTCATCGGAGTCTTGAAGAGGTTTACACCTCATCGCGGCTCA	720
RESULT 3	AX431809	Sequence 224 from Patent WO0229113.		Db	646	GTCTACTTGATGAACCTTTTCGACAGGTCAATTTGAAGCGCTTCACAGGTCAAAACGGGCA	705
LOCUS	AX431809	1958 bp DNA linear PAT 28-JUN-2002		QY	721	TTCCGATCAAGCTCATTTTCACGACGCGGAGGTATCTTGTCTTGTGGAATGAAAG	780
DEFINITION	AX431809			Db	706	TTTCTTTCATCTGTTTCTCTCGCGGCAACACCTTGCATTTATCGGAAACGAGAA	765
ACCESSION	AX431809			QY	781	GAATATAAGAAATGCTAGCTCTCAAGCGGTGGCTCTATGATATCGAACAAAGCGCGCTC	840
VERSION	AX431809.1	GI:21656613		Db	766	GAATATCAAAACGCGCCTCGATAAAGCATGGCTGTATGATATTTGAACAGGAAACCTA	825
KEYWORDS	Bacillus licheniformis.			QY	841	ACATGCTTACTGAGATGCTGGAGCTTCATTTACGGGATGCGCTGATTTGGAGATCAATG	900
SOURCE	Bacillus licheniformis			Db	826	ACGTGCTGACGGAATGCTTTGAGCTCCATCTCAGCGATGCGGTGGGACGAGACGCTT	885
ORGANISM	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.			QY	901	ATCCGTTGCTGACAGCGCGCGATTTTGGACAAAGGACACCGAGGCTTTTATGTCATC	960
REFERENCE	1	Berka, R. and Clausen, I.G.		Db	886	GTCGAGAGCTTCTTCCGAAACCGCGCTGGACCAAGACGAAACCGGCTTTTATGTCATC	945
AUTHORS	Methods for monitoring multiple gene expression			QY	961	GGCAGAGATCAAGGACGTACCGGCTCTATTATATTTTCGATTGAAGCGCTTTGTGATCCG	1020
TITLE	Patent: WO 0229113-A 224 11-APR-2002;			Db	946	GGATCTGACAGGCTTCTACAGGCTTTTACTACATATCAATCAAGGCTTTTGTCTTATCCG	1005
JOURNAL	Novozymes Biotech, Inc. (US) ; Novozymes A/S (DK)			QY	1021	ATTCGCTCGGAAAAAGAGTACATCAATAGCTTTTCTCTTTTACCTGATGAACAGACATTT	1080
FEATURES	Location/Qualifiers			Db	1006	GTCCGTTGGAAAAAGAGCATGTGAACGGAATTCAGCTTGCATCCGGACGAAAGTGGCTTT	1065
BASE COUNT	557 a 462 c 518 g 421 t			QY	1081	ATTCGAGTGTGACAAAGCCGACAGCCGAGTGAAGCTTTTACAGTATCC---CGCTTGA	1137
ORIGIN	1 ATGAAAAAGCTGATAACCGCAGACGACATCACAGCGATTGTCTGTGACCGATCCTCAA	60					

Query Match 41.08; Score 807.8; DB 6; Length 1958;
Best Local Similarity 64.3%; Pred. No. 1.3e-250;
Matches 1268; Conservative 0; Mismatches 687; Indels 18; Gaps 3;

Db 1066 GCGGCATCCATAGCCCTCGCCGTCGCGGCGAGCACTTTACCATCTCCAAATGGGTGAA 1125
QY 1138 CAGGAAGAGAAACAGCTACTCGCGCAATGACAAAGTTTGTTCAGGAGCATACGATATCA 1197
Db 1126 GAAAGAGCGGAGCTACAGATGCCAACAGGCTTTACGGAAGACATATCATATCG 1185
QY 1198 ATACCTGAGAGATCAATATATGCTACGAAGAGCGGCTGATGGTGAAGCGCTGCGTGTATG 1257
Db 1186 GAACCGGAGGAGCTCAATTTTCAGCAGCGGAGCTGTCGACAAATTCAGGATGGCTGATA 1245
QY 1258 AGGCTGACAAATGGAAGGTGAGACACATATCCACATTTCTTAAACATACACGCGGT 1317
Db 1246 AAGCGCCCAATATGAAAAAGAAACACGATATCGCTCAATCTGGAAGTCACGCGGGA 1305
QY 1318 CCGCATATGATGACGACATATATTTTCATGAGTTTCAGGCTCGCTGCGCGGCAAGGA 1377
Db 1306 CCTCAGCCATGATGCAATGCTTATTTTCATGAATTTCAAGTCTGCGCGCAAGGGA 1365
QY 1378 TAGCGGTGCTTTATATCAATCCGAGAGAGCCACGCTACGCGGCAAGAAATTTGTGAAT 1437
Db 1366 AGCGCGCTGATATGTTAAACCGCGGGAAGCCAGCGCTACGCGACGATTTGTCAAC 1425
QY 1438 CGGCTCAGACGAGATTTGGGGAAAGATTTATGACGATGTCGACGCTGTGTGATGAG 1497
Db 1426 CGGTGAGGCGGAGCTACGCGGTGGAGATTTTAAAGATGTAATGGCGCTGTGTGACAC 1485
QY 1498 GCTATCAACAGAGATCCGCATATGATCTTAAGCGGCTCGGTGTACGCGGCGGAGCTAC 1557
Db 1486 GTGCTGAACATTTATGTTCTGTCGATCAAGAAAGGCTCGCATCACCGGAGGCTAC 1545
QY 1558 GGAGGTTTATGACCACTGATCGCGGAGAGCAACCGCTTTAAAGCTGCGGTTACC 1617
Db 1546 GGGGCTTTATGACAACTGGGCTGTGCGGCACACGAAAGCGCTTCAAGCGCGCAGTGACA 1605
QY 1618 CAGCGCTCGATATCAAAATTTGATCAGCTTTTACGCGCTCAGTGATGCTATTTCTTT 1677
Db 1606 CAGAGGTCATTTCCACTGATCAGTTTACGGGTGAGGATGATGCTATTTTTC 1665
QY 1678 ACAGCTGGGAGCTGAGCATGATGATGTTTGGAGACACAGAAAGCTCTGGGACGGTCT 1737
Db 1666 ACAGCTGGGAGCTGCGGTGAGACCTTTTGAAGATCCAGCAACCTGTGGAGCATTCG 1725
QY 1738 CTTTAAATACGCAACGTCGAGACACCGCTTTTATGATGCTGCTGAAAGAAATGGGCAAG 1857
Db 1726 CCGCTCAATACGCGGACAGGTGGAGACCCCTTCTCATCTGCTGCGGCAAGCGGAC 1785
QY 1798 GACGATGCGGATGCGGAGCGGAGCAGCTGTTTATGCTGCTGAAAGAAATGGGCAAG 1857
Db 1786 GACAGTGGCGGATGAAACAGCGGAGACGCTGTTACGCGGCTGAAAGAAATGGGCAAG 1845
QY 1858 GAAACCAAGCTTGTCCGTTTTCGATGATGATGATGATGATGATGATGATGATGATGATG 1917
Db 1846 GAATGAACTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1905
QY 1918 AGACGCGGATCAAGCGGCTGATTTATGATGATGATGATGATGATGATGATGATGATGAT 1970
Db 1906 AAGCAGGATCAGCGGCTTGATATATGATGATGATGATGATGATGATGATGATGATGAT 1958

RESULT 4

CNSPAX06/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

265118 bp DNA linear BCT 09-SEP-1999
Pyrococcus abyssi complete genome; segment 6/6.

AJ248288 AL096836

AJ248288.1 GI:5458960

complete genome.

Pyrococcus abyssi.

Pyrococcus abyssi

Archaea; Euryarchaeota; Thermococci; Thermococcales;

Thermococcaceae; Pyrococcus.

1 (bases 1 to 265118)

Heilig, R.

Pyrococcus abyssi genome sequence: insights into archaeal

chromosome structure and evolution

unpublished

2 (bases 1 to 265118)

Genoscope.

Direct Submission

Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :

BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Location/Qualifiers

1..265118

/organism="Pyrococcus abyssi"

/strain="Orsay"

/db_xref="taxon:29292"

164..1615

/gene="bgal"

164..1615

/gene="bgal"

/note="PAB2376"

/codon_start=1

/transl_table=11

/product="BETA-GALACTOSIDASE (EC 3.2.1.23) (LACTASE)."

/protein_id="CAB50447.1"

/db_xref="GI:5458961"

/translation="MEFYGVVQSAFOFEMGDYPRNRDLTRSDMHWVDRDPVNIKNEL

VSGDLPEGINNVEYEDHDLAKELGYNAYOLAIENSRIFPCPIYSVEVERDGYG

FIKSVKIRKHELEKDELDELANKREVRHENVLNKLNFTTFVTLNHNQINPILWHDPI

EVRNIEKARGAWDERAIYVEFSKFAIYVAKEDKYVDFWATDEPMVTAELGYLAP

YVGNPPGLNPKAASVIMHQIIAARAYDSIKKPSDKPVGVLNIIIPAYPLNPRDSK

HVAKENYDLTFHNRLFLVANKGLDIDINGEYKVKHLKNDWIGNNYTYRVVYV

EPYKELPLTFVGVGYSANPNSIPDNNTSDFGWEYVYKGLYDSTAEAYEYSE

MVFTENGIDASKDILRPYIVDHVKVRLRENGINVLGYFHWALTDNYEWAGFKI

RFGIYVDPTIKERIPRRKSVEYTKVVKEGIE"

1834..2283

/note="PAB1019"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAB50448.1"

/db_xref="GI:5458962"

/translation="MPLPDDITFSDVTLKMHSEKRLLEITLAKTIVYLSMTIVOR

YLAATLKRNELEFEVLKDIYINVPVSEIILIRASKIANLRKIRKGFVLDLEIITA

VTAETSLITLDEPKYEPFIRRFGLDTPMLDKFLREVELMVEKEIT"

complement(2276..3079)

/note="PAB1341"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAB50449.1"

/db_xref="GI:5458963"

/translation="MFENLESFEFAFKLLEEVLEFDLNDPFDKVKVLCIEPHDDC

VIGMGITIKLTERGIEVIYACMTDGMGTLDSSLTGHTELATIRREERSSKLLGVK

KIYLVNRTDELPYSREVRKDLVIRIKKPDGVFLPDMPLPEAHDPDHRNTGFLAD

AVASPLNPSNDVDELGLGPHQVSYALYINKNPNFYVDITDMLKLAIRKTHKSO

FPDVMVEWEPFLRTVALFYKGRKAGVAGFGRIMPGLFYHITPFADLI"

3192..4340

/note="PAB1020"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

/protein_id="CAB50450.1"

/db_xref="GI:5458964"

/translation="MKEVSVSSYKEILVKLGLTDETRILEMKGIIEDFDFGIRYVR

FKDSAGLRRGTVIDEYVIFGPHKIRIINLRIRIFRKRGEYFVEKVGYNVR

VVMYKGLGITRGGICPFTTPEQFQFKNPNLILVGMAGPSPYLVGEP

PYVKEDIOFLFDQEIKTGRSLPVEERLKIAEYGINHVEYFGYTKDDVDLYOLI

ERLSKEGSIIMKSPDMKIKYVTPYANINDIKIGARFYELPGYFTRISRLAF

YLAERKIKGEERFAVEKLSALLQFSEFDVEQEDIEHLFKVVRKRIYAKMV

THFEKLGKIEIVDIEIKDWRITFRKRYLPDNTNIRELIGGKAFVD"

4351..5658

/note="PAB1021"

/codon_start=1

/transl_table=11

/product="hypothetical protein"

```
/protein_id="CAB50451.1"
/db_xref="GI:5458965"
/translation="MEDIALRVKAAIRAADPVKAVINSRVVGNKLLVVKQKEFEIKG
KYVLAFLGKAACDMARAEILIDVEDGVAVTKYGYKELKIRVIFAGHPIDENSIR
GAIEKLLEKVEEDIVFVLISGGGSALFELPEDGISLEDLKLINELLKSGAKIHE
INTVRKISKVGGKLAIKGTGIVLIIISDVVGNLEIAISGPTVKDPTTFQDAKRI
LELYGTDWVPESVSHIEKINGEAEETLKEELPNHNFILNASISCOAVPEEATK
LGFKAYMTTILEGADAGLFGISVIOGTAKTNPFPVPLVFGGETVTIIEGEGG
KGGPNQELASATRKIKGLNALIAFDTDGTDGPTDAAGIYDGETYERLRRKGIDIE
KVLKENSYEALKVGSLLFTGPTGTNVNSMIIAIIITSQKGT"
complement(5633..6655)
/notes="PAB1338"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB50452.1"
/db_xref="GI:5458966"
/translation="MPYIYKRVKVGISSSWVTSKVSRDALAOLESDRIYVKVLRMGE
VTGEDSFFSALDIDGIRIPDELKLPKXDFGLKYPVPCRGELLVLETFEENLLIYD
EKFAEFVHKVFEVLNGKTMVQLARIVGGALNMESKWEGERWIRVVKVKSARTQTE
RSVYVITQDRPVSIFSDLEIDIEIEVEMNGKVRWAKIRHFNHNSQVTSYLYIPDKK
VOLFVRLYLKIPSPAMEFIKIADDPFTLKSEFQELMEKELKESLDEMEKOKIIVA
LYSGINPLELHQFLGITEKEIEIEIYDRMIDKGLLKIWMIRKIVDLTNEGRKIYVNLK
YGLVSM"
complement(6670..7425)
/notes="PAB1337"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAB50453.1"
/db_xref="GI:5458967"
/translation="MDVGTLLILIAVVVLIGISVANSIVSSVISQMKMQLMISRLS
EVETKVEKKRTGSEAKPEEVAQEDVDSEEDIVRKIKQIDDEKRVHNAEAK
KERLLMDVARGALGFVSEDEYNAPLRLVISELDEFKRLWLAAYFPFEDKRLNQ
MLTHVARTKLPVIEVREKGGKESKLSPEALIRITSSKSAISLIDLLIKRQDNP
AVPLETEKLSQLEKKKVEDLEKRLQLEAL"
complement(7468..9687)
/notes="PAB1336"
/codon_start=1
/transl_table=1
/product="methyl-accepting chemotaxis protein (tlpC)"
/protein_id="CAB50454.1"
/db_xref="GI:5458968"
/translation="MOPRKIIILMAISAVITMAIFGVVTIYFNKVVGVKIGEQAP
AIEGGQLANSAEVAEKMEDDFARVADYGTMAETVVKYAVAGLDSNETAMREFLL
SEFKKIELNKDVSIVIFGDEGHMYPDELPEGDPRVRPVYQAKKEGGAIIIE
PYRDAATQWVIYAVPVYVNGQKVGILDFVSTLLNETKKIKLGTGIIAIDQN
GMTLVHPNEEPINKLNFQVPELKDADLKKDEGVLYTFQGIKKIGYKRLKT
GWIVATVPLSDITGPILSATDTAINKMKMVIQGFVGVIIIVAILVYIKASNTL
APLEKLYAAQALAEGLRKQVSEYKQIRYLERDEIGALIOAFVASKDLVGTLANAIS
KLERLAEGLDLSGLTYVEVGELEDITIEDIRSVTFPESIGSLVEMANDLEKRNAL
AQVSKDVTEALNQVNEALQQVSEIAQOQETINIEITDGMRLVAOTSESVRAMEFSG
AVTEVSTANSGSKQGDALKRIEDIOHMSKRIEETVSKVAESNRNIEETNVIYSIA
BOTNLLANAAIEARAGEAGRFVAVQAEIRKLAESKQADNIKSLIDKITDEIKE
AVEATKEGVISGESSETLRDTIGYLANIATLLQETSERMTTVKEQIVRTQEEVDKAL
RALENLAASRETTASAEVSSAIEQOQTAATEELRRAAQELKDMVGRMRQIVGKFKV"
complement(9704..10195)
/notes="PAB1335"
/codon_start=1
/transl_table=1
/product="CHEMOTAXIS PROTEIN CHED"
/protein_id="CAB50455.1"
/db_xref="GI:5458969"
/translation="MTRKIKVIGDYAVNGEGEIIISTYGLSCVGLIYDRVTKVGGI
LHALPEARVYGHGRGNPKARYVDIGLQLLKEVLKGLASKFRLEAKLFGAQNQNIKS
DELKTEGRNVQATKRELKGLIRLVAEDTGGRGRTIYLDLSTGKVRMRKVIGGVIE
KVI"
complement(10179..10817)
/notes="PAB1334"
/codon_start=1
/transl_table=1
/product="CHEA HISTIDINE KINASE"
/protein_id="CAB50456.1"
```

```
/db_xref="GI:5458970"
/translation="MKKSELYKDFIKFASNIAMSHALTSLSOMIGPIEMEADPDEVIV
SRAEFKLKLAERGVSKGFTVMFDVTEGLSGLTIIQFPKHSALNLSAVLMGERGSHEE
LDMGKSAIMEVGNILISAYDILSNLIGEPSLSPKPAESLYDIEKEGRDPLRNV
NSIWFKSPFYKEDIGVESFYILVPTPESPSTKLKVKLEQKISEVEANDBN"
complement(10814..11452)
/notes="PAB1333"
/codon_start=1
/transl_table=1
/product="CHEA HISTIDINE KINASE related protein"
/protein_id="CAB50457.1"
/db_xref="GI:5458971"
/translation="MDIAENIEEYVKNLDEFKSALETFTNIGASHAATALSMTGR
EVNLSVPLKAVEIKNVEIVGEDVKVYIELGKDFSHAFPAVDYDVLDFMDVIM
GNPPTITELDDMKVSKSFMENGLISAFANALSQELGITIEQTPPALTIIDFLPAILD
LALADIGCDFDITIIQDIIKISGVFEFEEHLLFPKPEDMKKILEKLLGLLA"
complement(11436..13754)
/notes="PAB1332"
/codon_start=1
/transl_table=1
/product="chemotaxis histidine kinase (cheA)"
/protein_id="CAB50458.1"
/db_xref="GI:5458972"
/translation="MVTLEMSQYLDLADARIDRISLSNAITLEKIVKEGSEEE
KAELINQIFRDAHTLKGTAATMGFMKLESETAHKMFDAIRNSQIELTPEVDLVLD
FLDAIEAMVNNIEESGEGDIDSDLFRRADFLQGGKGAEKPEEKPEEVEAKEE
EKPEVEGVGVYHLKVVQKAPLKVRAFLISDLERGVFWITNPGRDVIESG
NLDKDVLEFKVSSDLPEELEKVIIRHPEVEKVEIEEGEEKGPEGEKVELYLRVYF
SPDAPLKPSYLILODLESIGDIVQTNPNRGDIENGLLGKFPFVYLRKEDKSKI
VOLIKHPDVENFVYTAEEAEKEKKEKPPQKPPKVPVETPRVKISRII
KIDVGHLDKLMNLVGLVITKLEQIEGRGLDELFALSTLSRLTELODEIMEMR
LTPAEVFNKPRMVDLARKKVEVEFMEGADIEVDRTILEKIGDALVHLLRNADV
HGIPEERVKLGRVGRVELIAKREKNHVVIIVRDDRGIDPEKVKKAKIERGLIT
PEEASLDEEAINLIELPFSTAEKVKVMDVVKVSKMGSISVQTEVG
KGTIFILKPLSMIAIOLAIRVMDETYAIPNNILETIEVDPSTLTKTGKPVILVR
GEIIPVMLHELFGIPTPQVPERFALIVDHGAQKVAIGVDELLHRRDIVKSLGKFLS
NVRFGATILGDGSVLIIDIGFLLGGGYSG"
complement(13717..14823)
/notes="PAB1331"
/codon_start=1
/transl_table=1
/product="protein-glutamate methyltransferase (cheB)"
```

```
Query Match 12.1%; Score 238.2; DB 1; Length 265118;
Best Local Similarity 55.3%; Pred. No. 1.4e-64;
Matches 482; Conservative 0; Mismatches 386; Indels 3; Gaps 1;

QY 1091 TGACAAACGGGACAGACCGGAGTGTACAGATATCCGCTTGACAGAGGAGAGAAAC 1150
DB 58851 TTACAGCTCAAGATCGGACAACTCCGACTGAGCTTTACCGCTTCAGGGATGAAAGAGA 58792
QY 1151 AGCTCACTGGCGCGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATACCTGAAGAGA 1210
DB 58791 ACCGTATCACAGACTTTACGCGCTGGATAGGACTATAAATTTTCAAAACACAGACACT 58732
QY 1211 TTCAATATGCTACAGAGACGGCGTGATGGTGAACGGCTGGCTGATGAGCCCTGCACAA 1270
DB 58731 TCAGGTTACGGCTAGCGATGGGAGAAATAGATCGCTGGATCATGAAGCCCGTTGACT 58672
QY 1271 TGAAGGTGAGACACATATCCACTTATTTCTTAACATACACGCGCTCCGATATGATGT 1330
DB 58671 TTAAGCCTGGAAGGAGATCTCTGAGTCTGCAATTCACGGTGGGCCAAACACGCT 58612
QY 1331 ACCGACATACATATTTTCATGAGTTTTCAGTGTCTGGCGGCGAAAGATACGCGGTGCTTT 1390
DB 58611 ACGGTATGCAATTTATGCAGAGTTTCAGGTTTTTACTTGCCTCAAGAGCTTCGTTGTAATCT 58552
QY 1391 ATATCAATCCGAGAGGAGCCAGCGCTACGGCAGAAATTTGTGAATCGGCTCAGAGAG 1450
DB 58551 TTTCAATCCAGGAGGAGTATGTTTATGGTAGGAGTTCGCG---GACATANGGGTC 58495
QY 1451 ATTATGGGAGGAGGATTTATGACGATGTGATGACGCTGTGGATGAGGCTATCAACAGAG 1510
```


Db 58494 ACTACGGTGAGCGGATTAATCAAGACCTAATGAGGTAGTTGACGAGGCTTTAAACGCT 58435

QY 1511 ATCCGATATTGATCTAAGCGGCTCGGTGTCACGGCGGAAGCTACGAGGTTTATGA 1570

Db 58434 TTGACTTTATAGATGAGGAAGGCTTGGCTTACCGAGGATCATACGAGGCTTTCATGA 58375

QY 1571 CCAACTGGATCGTGGGAGAGCAACCGCTTTAAAGCTGCCGTTTACCGAGGCTCGATAT 1630

Db 58374 CAAATTTGGATCTGTTGGGACACAAACAGATTCAAAGCTGCCGTAACCTCAGCTCAATAT 58315

QY 1631 CAAATTTGGATCGTTCACGCGCTCAGTGATATCGGCTATTTTACAGACTGGCAGC 1690

Db 58314 CCAACTGGATCGTTCCTTGGGACGACCAACAGATTCAAAGCTGCCGTAACCTCAGCTCAATAT 58255

QY 1691 TTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGGCTCTTTAAATACG 1750

Db 58254 TCGGTGGGATCTTGGAGTAACCTAGAGGCTACTGGGAGAGAGCCATTTGAAGTACG 58195

QY 1751 CAGCAAACTGGAGACCGCTTTTGTACTGCGATGCGGAGGATGACCGATGCGCGA 1810

Db 58194 CTCCAAAGCTTGAGACTCCGCTCTGATAATTCACCTCCAGGAGGACTACCGCTGTTGGC 58135

QY 1811 TCGAGCAGCGGAGGAGCTGTTTATCGCTCTGAAAAAATGGGCAAGGAAACCAAGCTTG 1870

Db 58134 TTCCCGAGGCTTTGCAATTTTACATAGCTTTGAAGTATTTGGGCAAAACCCCTCAGCTGG 58075

QY 1871 TCCGTTTCCGATGATCGCAATTTTATCAGCAGCCGACACCCAGAGAGCGGATCA 1930

Db 58074 CGATATTTCCAGGTGAGATCATGACTTGTAGTAGGAGTGAAGGCCAACCATAGGCTTA 58015

QY 1931 AGCGCTGAATTTATCATGCTCATGTTTGA 1961

Db 58014 AGAGGCTCGAGCTCATAGTGGGTGGTTGA 57984

RESULT 5
AX041922/c

LOCUS 265118 bp DNA linear PAT 23-NOV-2000

DEFINITION Sequence 817 from Patent WO0605062.

ACCESSION AX041922

VERSION AX041922.1 GI:11340687

KEYWORDS Pyrococcus abyssi.

SOURCE Pyrococcus abyssi.

ORGANISM Pyrococcus abyssi.

REFERENCE 1 (bases 1 to 265118)
Archaea; Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
Porter, P., Thierry, J.C., Prieur, D., Dietrich, J., Lecompte, O.,
Querrelou, J., Weissenbach, J., Saurin, W., Heilig, R., Flament, D.,
Raffin, J.P., Henneke, G., Gueguen, Y. and Rolland, J.L.
Genome sequence and polypeptides of Pyrococcus _i(abissy), fragment
and uses thereof
Patent: WO 0605062-A 817 02-NOV-2000;
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE (FR) ; IFREMER
INSTITUT FRANCAIS DE RECHERCHE POUR L'EXPLOITATI; ON DE LA MER (FR)
Location/Qualifiers
1..265118
/organism="Pyrococcus abyssi"
/db_xref="taxon:29292"

FEATURES
source
/note="Original length of seq 1: 1765118 (for 812
sequences) New sequences : seq 001, from 0.000.001 to
0.349.980, length: 349.980 seq 813, from 0.300.001 to
0.649.980, length: 349.980 seq 814, from 0.600.001 to
0.949.980, length: 349.980 seq 815, from 0.900.001 to
1.249.980, length: 349.980 seq 816, from 1.200.001 to
1.549.980, length: 349.980 seq 817, from 1.500.001 to
1.765.118, length: 265.118"

BASE COUNT 75226 a 61302 c 54005 g 74585 t

ORIGIN

Query Match 12.1%; Score 238.2; DB 6; Length 265118;
Best Local Similarity 55.3%; Pred. NO. 1.4e-64;
Matches 482; Conservative 0; Mismatches 386; Indels 3; Gaps 1;

RESULT 6
AR116209

LOCUS 1896 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 2 from patent US 6133012.

ACCESSION AR116209

VERSION AR116209.1 GI:14096531

KEYWORDS Unknown.

SOURCE Unknown.

ORGANISM Unclassified.

REFERENCE 1 (bases 1 to 1896)

QY 1091 TGCACAAAGCCGACAGACGAGTACGCTTTACAGTATCCGCTTGGACAGGAAGAAAC 1150

Db 58851 TTACAGCTCAAGATCGGACCACTCCGACTGAGCTTTACGCTCTCAGGATGGAAGAGA 58792

QY 1151 ACCTGACTGGCCGAATGACAAAGTTTGTTCAGGAGCATACGATATCAATACCTGAAGAGA 1210

Db 58791 ACGGTATCACAGACATTTACGCGCTGATTTAGGAGCTATAAATTTTCAAAACACGAGCACT 58732

QY 1211 TTCAATATGCTACAGAAGACGCGCTGATGTTCAAGCGCTGGCTGATGAGCGCTGCACAAA 1270

Db 58731 TCAGGTTTACGCTAGCGATGGAAGAAATAGATGCTTGGATCATGAAGCCGCTGACT 58672

QY 1271 TGGAGGTGAGACACATATCCACTTATTTTAAACATACACGCGGTCGCGATATGATG 1330

Db 58671 TTAAGCCCTGGGAAGGAAGTATCTCGAGTGTGGAATTCACGCTGGGCAAAACAGCCT 58612

QY 1331 ACGGATACATATTTTTCATGAGTTTTCAGGCTCTGCGCGGAAAGATACGCGTCTGTTT 1390

Db 58611 ACGGTTATGCATTTATGCACGAGTTTTCAGGTTTATTTAGTGGCCAAAGGCTTCGTTGTAATCT 58552

QY 1391 ATATCAATCCGAGAGGAACCGCTACGGCAGGAATTTGTGAATGCGTTCAGAGGAG 1450

Db 58551 TTTCAATCCAAAGGGAAGTATGTTTATGTTGAGGAGTTCGCG--GACATAAGGGGTC 58495

QY 1451 ATTATGGGGGAAAGGATTTATGACGATGATGTCAGGCTGTGGATGAGGCTATCAACGAG 1510

Db 58494 ACTAGGTGAGCGGATTTATCAAGACCTAATGGAGGTAGTTGACGAGGCTTTAAACGCT 58435

QY 1511 ATCCGCATTTATGCTTAAGCGGCTCGGTGTCAGGGCGGAAGCTTACGGAGGTTTATGA 1570

Db 58434 TTGACTTTATAGATGAGGAAGGCTTGGCGTTACCGGAGGATCATACGAGGCTTCATGA 58375

QY 1571 CCAACTGGATCGTGGGAGAGCAACCGCTTTAAAGCTGCGGTTACCCAGGCTCGATAT 1630

Db 58374 CAAATTTGGATCGTTGGGACACAAACAGATTCAAAGCTTCGCGTAACCTCAGCGCTCAATAT 58315

QY 1631 CAAATTTGGATCGTTCACGCGCTCAGTGATATCGGCTATTTTACAGACTGGCAGC 1690

Db 58314 CCAACTGGATCGTTCCTTGGGAGCGGATATCGGCTATTTACTTCGCTCCAGATCAGA 58255

QY 1691 TTGAGCATGACATGTTTGGAGACACAGAAAGCTCTGGACCGGCTCTTTAAATACG 1750

Db 58254 TCGGTGGGATCTTGGAGTAACCTAGAGGCTACTGGGAGAGAGCCATTTGAAGTACG 58195

QY 1751 CAGCAAACTGGAGACACCGCTTTTGTACTGCGATGCGGAGGATGACCGATGCGCGA 1810

Db 58194 CTCCAAAGCTTGAGACTCCGCTCTGATAATTCACCTCCAGGAGGACTACCGCTGTTGGC 58135

QY 1811 TCGAGCAGCGGAGGAGCTGTTTATCGCTCTGAAAAAATGGGCAAGGAAACCAAGCTTG 1870

Db 58134 TTCCCGAGGCTTTGCAATTTTACATAGCTTTGAAGTATTTGGGCAAAACCCCTCAGCTGG 58075

QY 1871 TCCGTTTCCGATGATCGCAATTTTATCAGCAGCCGACACCCAGAGAGCGGATCA 1930

Db 58074 CGATATTTCCAGGTGAGATCATGACTTGTAGTAGGAGTGAAGGCCAACCATAGGCTTA 58015

QY 1931 AGCGCTGAATTTATCATGCTCATGTTTGA 1961

Db 58014 AGAGGCTCGAGCTCATAGTGGGTGGTTGA 57984


```
/transl_table=11
/product="120aa long hypothetical protein"
/protein_id="BAA29696.1"
/db_xref="GI:3257013"
/transl_table=1
KSNLNVGGRSICSDPLFILLNLPCGFAAITVVLAFSVAVMPAPDRILVCCSRAS
WITLLSLILSSSSINA"
3592..3981
/gene="PH0608"
CDS
3592..3981
/gene="PH0608"
/codon_start=1
/transl_table=11
/product="125aa long hypothetical protein"
/protein_id="BAA29697.1"
/db_xref="GI:3257014"
/transl_table=1
GTSLSFWITLRIPTTITLSPGIAVSTISSRSIPSMNLGICPPLKLSGLCILNIWKF
TNLFSSTWTFLLPHCSHFFATKPS"
3693..4040
/gene="PH0609"
CDS
3693..4040
/gene="PH0609"
/codon_start=1
/transl_table=11
/product="115aa long hypothetical protein"
/protein_id="BAA29698.1"
/db_xref="GI:3257015"
/transl_table=1
IYPRHLPPSSQALWPILYSEHVEFALLHIKIDVFTSALLAPLYCNKALIGLLRH
HFWPAITNEDRD"
complement(4439..4936)
/gene="PH0610"
complement(4439..4936)
/gene="PH0610"
/note="motif=DEAD and DEAH box families ATP-dependent
helicases signatures"
/codon_start=1
/transl_table=11
/product="165aa long hypothetical protein"
/protein_id="BAA29699.1"
/db_xref="GI:3257016"
/transl_table=1
IIDEADVIGEIDKVLREKRNAPISALPENLYRAEFYIVLEPKYIQKGEGNIDVI
HTKLTGLKNLKKYKMLKDIRFKKILEAVRLRPNSEILSLAPQERRIYLQISIRN
EWIGE"
complement(4979..5737)
/gene="PH0611"
complement(4979..5737)
/gene="PH0611"
/note="similar to Swiss_Prot:P46577 percent
identity:30.736 in 238aa. motif=ATP/GTP-binding site motif
A (P-loop)"
/codon_start=1
/transl_table=11
/product="252aa long hypothetical protein"
/protein_id="BAA29700.1"
/db_xref="GI:3257017"
/transl_table=1
MGGTMIIVFGTAGSGKTSLTAFGEVLEENYKAVYNLDTGVK
ELPKYDIDREFVVEIKMEKGYGPNAGVIESYDLRMNKFYLYLKEILELESKKDY
LIDTPGQMTFLPHEFGVKLMENLPYLVVYLSDPTEILRKPTDYCFVRFALLIDLRL
GATVPALNKVDLLKEOLEKHRYFEDLDYLMGLKFPDPSMQGLMAYKMCISIMNEVL
PPIRVLYLSAKTRREGFDELTLAYEHYCTGDLI"
complement(5768..6505)
/gene="PH0612"
complement(5768..6505)
/gene="PH0612"
/note="similar to PIR:C64368 percent identity: 41.379 in
233aa; PIR:B64321 percent identity: 35.841 in 227aa;
owl:S76970 percent identity: 34.188 in 239aa.
motif=ATP/GTP-binding site motif A (P-loop)"
/codon_start=1
/transl_table=11
```

```
/product="245aa long hypothetical cell division inhibitor
Mind"
/protein_id="BAA29701.1"
/db_xref="GI:3257018"
/transl_table=1
NLSVLGVDDVNIITLHDVLADAKLEDAIYMTQENVYILPCAVDWEHVIKADPRKLP
EVKSLGKYDFILIDCPAGLQDAMSAMLSSEALLVNPETISCLTDTMKVGMVLKK
AGLAILFILNRYGRSERDIPPEAAQDVMYDPLLAIVIPEDPVIREGTLEGIPAVKYKP
ESKGAQAFIKLAEEVDKLAGIKAKIMY"
Query Match 11.5%; Score 227.2; DB 1; Length 233000;
Best Local Similarity 55.4%; Pred. No. 5.2e-61;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;
Qy 1139 AGAAGAGAAACAGCTGACTGGCGGAATGACAACTTTGTCAGGAGCATACGATATCAA 1198
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228212 ATGAAAGAGAGAAGGTTACCGACTTTTAAACAATGATAAAGGTTACACCCTTCAA 228271
Qy 1199 TACCTGAAGAGATTCAATATGCTACAGAAGACGGCGTGATGGTGAACGCTGGCTGATCA 1258
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228272 AACCTGAACACTTTAAGGTTAAAGCAAGTGACGGGTTGAATAGATGCTGGGTAATGA 228331
Qy 1259 GGCCTGCACAAATGGAAGGTGAGACACATATCCACTTATTCTTAAACATACACGCGGTC 1318
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228332 AACCGGTGAACCTTCAGGAAGAAAGAAAGTATCCAGCTATTCTACAGATCCACGTTGGTC 228391
Qy 1319 CGCATATGATGTAGGACATACATATTTTCATGAGTTTCAGTGTCTGGCGGGAAGGAT 1378
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228392 CTAACACCGCTTACGGTTTACGCTTTTATGCACGAGTTCCACGTTTAAACCTCTAAAGCT 228451
Qy 1379 ACGCGTCTGTTTATATCAATCCGAGGAAGACCGCTACGGGCAAGAAATTTGTGAATG 1438
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228452 TCGTCGTGATATTCTCAATCCTAGAGGAGCGATGCGTACGAGAGGAGTTCCGCGAT- 228510
Qy 1439 CGGTCAGAGAGATATTGGGGAAGAAAGATTATGACGATGTGATGACGCTGTGGATGAGG 1498
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228511 --ATAAGGGGACACTATATGGGAGAGGATTACACAGGATTTAATGGAGTAGTCGATGAAG 228568
Qy 1499 CTATCAACAGAGATCCGCATATTGATCTTAAGCGGCTCGGTCTACGGGCGGAAGCTAGC 1558
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228569 CATTAAAGAGATTTGACTTCATAGATGGGAAAGCTAGAGTTACCGGGGTTCTTATG 228628
Qy 1559 GAGGTTTTATGACCAACTGGATCGTCGGGCGACACAAACCGCTTTAAAGCTGCGGTACCC 1618
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228629 GTGGCTTCATGACGAACCTGGATAGTCGGACATACCAACAGGTTCAAGCCGCTGTAACCC 228688
Qy 1619 AGCGCTCGATATCAAAATTTGGATCAGCTTTTCAGGCGGTGATGATTCGCTATTTCTTTA 1678
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228689 AGAGATCAATTTCAAAATTTGATAAGCTTCTTCGGGACACACGATATAGTTTACTTTG 228748
Qy 1679 CAGACTGCGAGCTTCAGCATGACATGTTTGAGGACACAGAAAGCTCTGGGACCGCTCTC 1738
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228749 CTCAGATCAAAATGAAAGAGATCCCTGGAGCAACTTGAAGGTTATTGGGAAAGAGGCC 228808
Qy 1739 CTTTAAATACGCAACACGTTGGAGACACCGCTTTTGATCTGCTGCGAGCGGATG 1798
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228809 CATTAAAGTACGCTCCCAACGTTGAACCTCCCTGCTTATATATCCACTTACCGAAGACT 228868
Qy 1799 ACCGATGCCGATCAGAGAGCGGAGAGCTGTTTATCGCTCTGAAAAAATGGGCAAGG 1858
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228869 ACAGGTGTGGCTTCCCGAGGCAATTCGCACTCTTCATATCCCTAAATACCTGGGGAAGA 228928
Qy 1859 AAACCAAGCTTCTCGGTTTTCCGAATGATCCCAATTTATCAGCAGCCGACACCCAA 1918
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228929 GAGTTGAATTTGGCAATATTCCAGAGAGAAATCATGACCTAAGTAGATCTGGGAGGCCAA 228988
Qy 1919 GACAGCGGATCAAGCGCTGAAATTTATCAGCTCATGTTTGTGATCAACATCT 1970
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
228989 AGCAGGTTTAAAGACTTGAACATAATAGCAGGATGGATGGAGAAATGGCT 229040
RESULT 9
AE013176/c
```

LOCUS AE013176 12882 bp DNA linear BCT 09-MAY-2002
DEFINITION Thermoanaerobacter tengcongensis strain MB4T, section 203 of 244 of the complete genome.
ACCESSION AE013176 AE008691
VERSION AE013176.1 GI:20517339
KEYWORDS
SOURCE Thermoanaerobacter tengcongensis.
ORGANISM Thermoanaerobacter tengcongensis
Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
Thermoanaerobacteraceae; Thermoanaerobacter.
REFERENCE
AUTHORS Bao, Q., Tian, Y., Li, W., Xu, Z., Xuan, Z., Hu, S., Dong, W., Yang, J., Chen, Y., Lai, X., Huang, L., Dong, X., Ma, Y., Ling, L., Tan, H., Chen, R., Wang, J., Yu, J., and Yang, H.
TITLE A Complete Sequence of the T. tengcongensis Genome
JOURNAL Genome Res. 12 (5), 689-700 (2002)
MEDLINE 21992816
PUBMED 11997336
REFERENCE
AUTHORS Bao, Q., Xu, Z., Hu, S., Dong, W., Chen, Y., Wang, J., Yu, J., and Yang, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Beijing Genomics Institute/Genomics and Bioinformatics Center, Institute of Genetics and Development, Chinese Academy of Sciences, Beijing Airport Industrial Zone B6, Beijing 101300, China
REFERENCE
AUTHORS Li, W., Xuan, Z., Yang, J., Ling, L., and Chen, R.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, China
REFERENCE
AUTHORS Tian, Y., Xue, Y., Xu, Y., Lai, X., Huang, L., Dong, X., Ma, Y., and Tan, H.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2001) Institute of Microbiology, Chinese Academy of Sciences, Beijing 100080, China
FEATURES
source
1. 12882
/organism="Thermoanaerobacter tengcongensis"
/strain="WB4T"
/db_xref="taxon:119072"
terminator complement(70..90)
gene complement(93..1765)
/gene="TTE2335"
CDS complement(93..1754)
/gene="TTE2335"
/note="Best Blastp hit = gi|10175658|dbj|BAB06755.1| (AF001517) two-component sensor histidine kinase [Bacillus halodurans], score 177, E-value 4.00E-43"
/codon_start=1
/transl_table=11
/product="Diguanylate cyclase/phosphodiesterase domain 1 (GGDEF)"
/protein_id="AA025476.1"
/db_xref="GI:20517340"
/translation="MGKRYELFNLTALATGGLVFAPALXHYGLGIDFLRLILLFLA IILDNGIMTYDKLSLSPGSGVIAAFLFGTVGAATLMTVSMVDPTLVMRRKIKNGFL NGAMFALYLPACGLVPEFMGKIGIEISIAQVKYILVVIVISFLNNLTLYALAVQGL ILPEKWNESVLLEGTIVLMPVPAAMLLAYPKHSLMFEVLSLTPLVFAYVFMVYR DLVANKRLNLAIEYMKMINSKLELDQILTIVETISQVAVSAAALYLTDSNGIATL KVAGKGRGKRESFKGEGGLIGKVISSNKIVAINKEDRSIDREKIFANYGSLIA APLRSSGVPICGLLIHLHETNAFDSDSRIIEIIVDQASVIAITNAKKYIEVTRKSID PLTKTYNRRYFDALMENIMRADENEPVSLIMFDLNFQKINDTYGHLIGDEVLKEV ARRIKNNRNDIVARFGGEFAVILPKLTAEOAYMIAERIRTVSSKPIKTEKGDIY VTITGVADYPSKADSAEKLVSHADRALYAGGSKRGNKTAIEV"
misc_feature complement(114..611)
/gene="TTE2335"
/note="Pfam match to entry GGDEF, GGDEF domain, score 212, E-value 9.30E-60"
misc_feature complement(612..1031)
/gene="TTE2335"
/note="Pfam match to entry GAF, GAF domain, score 59.7, E-value 6.40E-14"

RBS complement(1761..1765)
/gene="TTE2335"
terminator complement(1912..1934)
/note="putative rho-independent transcription terminator"
gene complement(1941..4016)
/gene="NrdA"
/note="TTE2336"
CDS complement(1941..4016)
/gene="NrdA"
/note="Best Blastp hit = gi|11499254|ref|NP_070492.1| ribonucleotide reductase (nrd) [Archaeoglobus fulgidus] gi|17483938|pir|G69457 ribonucleoside-diphosphate reductase (EC 1.17.4.1) large chain nrd - Archaeoglobus fulgidus gi|2648891|gb|AA89584.1| (AE000988) ribonucleotide reductase (nrd) [Archaeoglobus fulgidus], score 440, E-value 1.00E-122"
/codon_start=1
/transl_table=11
/product="Ribonucleotide reductase alpha subunit"
/protein_id="AA025477.1"
/db_xref="GI:20517341"
/translation="MFKRVAKAIATKEEEMFYKVMSEWKFIPGGRILAGAGTGRGV TYNCFVIVPEANDPKGNSDRAAIMDTIAKMLEINSRGGGVGNWTLRPRGAYVKV VGGTSSGAYWMLAANEVATQVGGSRRAALMFLMDWHPDIEEFIKVKDLTKMQQ ANLSVADAFMEAVDAQKWKLEFPDTSHPNDVDAEWKGLNEWKAGYPTVYVKEIP ARELNKIVTAAWESAEPGVVFLERYNKLSTVYITKIIITNPGELGLEPYGVCNLG AINLVAFVNGKINFEELETVKIAVRFLDNVLDLGYVLPQNKEMAQKLRVGLGLM GLADAILMNRYSSEESUKATEEIVRRIRDAAYRASVELAKEKGPPEFVADKYLKG QTIQRPKDIRDEIAKIRNATILTQAPTGTTSILAGSSGIEPNPAKEYVRKDRGT THTVHKLADYPAFVAHEVTPPEHVQMVAQLOKYIDSSISKTINLPKATVDDIDNI YLITFDNKPFEFVNIAGRSVDAFTAIALRLISLALRGVAVEDIIEQLVGIGGA TTIGGGRNVLSPDGAIGKALNRNGYEQGNSVKVSATNVDICPCVGMATFYTEGCKT CLNCGYSQC"
misc_feature complement(2442..3872)
/gene="NrdA"
/note="Pfam match to entry ribonuc_red.lgC, Ribonucleotide reductase, barrel domain, score 126.6, E-value 4.60E-34"
gene complement(4033..4224)
/gene="TTE2337"
CDS complement(4033..4209)
/gene="TTE2337"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AA025478.1"
/db_xref="GI:20517342"
/translation="MHLKLEKNTKCIIEDVDIONIVFYVTKGKINASSVERTEGEGF SQGVCLKRGQGR"
RBS complement(4220..4224)
gene complement(4297..4808)
CDS complement(4297..4794)
/gene="TTE2338"
/note="Best Blastp hit = gi|11349142|pir||A83493 hypothetical protein PA1234 [imported] - Pseudomonas aeruginosa (strain PA01) gi|1947163|gb|AG04623.1|AE004552.13 (AE004552) hypothetical protein [Pseudomonas aeruginosa], score 55.8, E-value 3.00E-07"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA025479.1"
/db_xref="GI:20517343"
/translation="MSLKDLFEKGLTFDEFIETSDITRQIRERYDKTVLEKDVQK KIVSLPKVNIILAFATETPCDSQVNPVIAKIASYNKTFELKIVKREGNEEYMKPYVE GKPRIPTFVMDENEFKFWHVERPOIVRELEKKEVADWKRDLKGYDSEIIRIIVY IFLRP"
RBS complement(4804..4808)
gene complement(4809..6244)


```
QY 1559 GAGGTTTATGACCAACTGGATCGTGGGCGAGACGCGTTTAAAGCTGCGGTACCC 1618
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 67 GTGGCTTATGACCAACTGGCTGTGGCCATAGCAATCGGTTCAAGCGCGGTATCGC 126
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1619 AGCGCTCGATCAAAATGGATGACGCTTTCACGCGTCAGRATATCGCTATTTCTTTA 1678
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 127 AACGCTCGATATCAATGGATTAGTTTATGCGGTAAGCGATATTCGCTACTATTTT 186
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1679 CAGACTGGCAGCTTGACATGACATGTTTGAGGACACAGAAAGCTCGGACCGGCTC 1738
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 187 CAGAGTGGCAATATAAGACAGAC---TTGCATGATATCGAAACACTGTGGGCCCATTCGC 243
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1739 CTTTAAATACGCGACGAAACGTCGAGACACCGCTTTTGATATCTGATCGCGCGGATG 1798
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 CACTCAATAGCTTGAACAGCTGGAAGCGCAATTCGCTTATTCGCACGCTGAAAGAAAT 303
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1799 ACCGATCCGATCGAGCAGCGGAGCGCTGTTTATCGCTCTGAAAAAATGGCAAGG 1858
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 ATCGCTCGGATTTGAACAGCAGCAGCAATTTGTTTATTCGCTTGAAGAAAGCATGGGAAG 363
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1859 AAACCAAGCTTGCTCCGTTTCCGATCGATCGGACAAATTTATCAGCACCGGACACCCAA 1918
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 364 AGACTGTATTATCTCCGTTTCCGAGGCAATCATGAACTGTCAGAAAGTGGTAAACCTA 423
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1919 GACAGCGGATCAAGCGCTGAAATATATATCAGCTCATGCTT 1959
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 424 ATTTCGGATGACGCTCAATCGATAGCAGATTGGTTT 464
  ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
AE001879/c
LOCUS
DEFINITION
  Deinococcus radiodurans RI section 16 of 229 of the complete
  chromosome 1.
ACCESSION
  AE001879 AE000513
VERSION
  AE001879.1 GI:6457832
KEYWORDS
  .
SOURCE
  Deinococcus radiodurans.
  Bacteria; Thermus/Deinococcus group; Deinococci; Deinococcales;
  Deinococcaceae; Deinococcus.
  1 (bases 1 to 10256)
  White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
  Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
  Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
  Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
  Makarova, K.S., Aravind, L., Daly, M.J., Fraser, C.M. et al.
  Genome sequence of the radioresistant bacterium Deinococcus
  radiodurans RI
  Science 286 (5444), 1571-1577 (1999)
  20036896
  10567266
  2 (bases 1 to 10256)
  White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
  Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
  Moffat, K.S., Qin, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
  Vamathevan, J.J., Lam, P., McDonald, L., Utterback, T., Zalewski, C.,
  Makarova, K.S., Aravind, L., Daly, M.J., Minton, K.W.,
  Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
  Smith, H.O., Venter, J.C. and Fraser, C.M.
  Direct Submission
  Submitted (08-NOV-1999) The Institute for Genomic Research, 9712
  Medical Center Dr, Rockville, MD 20850, USA
  Location/Qualifiers
    1..10256
      /organism="Deinococcus radiodurans"
      /strain="RI"
      /db_xref="taxon:1299"
      /chromosome="1"
      100..1017
        /gene="DR0164"
      100..1017
        /gene="DR0164"
      /gene="DR0164"

FEATURES
Source
  gene
  CDS
  /note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776
  GB:X68777 percent identity: 62.16; identified by sequence
  similarity; putative"
```

```
/note="similar to GB:D14846 PID:286165 percent identity:
49.11; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="endo alpha-1,4 polygalactosaminidase, putative"
/protein_id="AAF09753.1"
/db_xref="GI:6457833"
/translation="MRPLTFLVLLPLSAVLASCGTSPVAATVFPATPAAPRTTLA
AQATPLRPPAGKLAQWQIGAEAKVALPAGVSLDLDDGFTSAKVALDLKAQGV
YVYLVNVSYESYRDAAYPDLSKIQTDPNMPDESFDVIRDFRGGVLAGILDRR
LALCAAKGDAVEPDNDQNVTSVSRQDQDFNGWLADRAHAGLA TLQKNGPD
VYLOADRQGLMVDLFDGVLNESCQRYKEGGLTEYVVRGKLALNVEYRQADLNCATM
NSLGVNALFKDLISLVGKLQKAYKRVSCG"
complement(1079..3046)
/gene="DR0165"
complement(1079..3046)
/gene="DR0165"
/note="similar to GB:AL009126 percent identity: 56.01;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-peptide hydrolase, putative"
/protein_id="AAF09754.1"
/db_xref="GI:6457834"
/translation="MNNSETPAPGPDLSLLALAFSPQVSPDGCKQVAFVLAQISEEDP
AKPDKDFAPRYSGLWSEGGAAARPLTHAETGSGAPRWSPDQNLAFVRSAGEVK
AALMLPLKGGEARVTHFKNGVSPQSDGRIATFTATTEDTKRDERGEARVLTR
PYRANGADMLPAPAAWLIVYADKLRWYAPETIGALSWMPSDGRVLIVGSEDE
WQASQWRQVYDPLPLTADAPAPAKLLDWNAAHGLADHPDQRFALIGRPAGKNGT
EHAHLYLIENGHRRLDTHDHPGDAVGDCGHVAFPEGPRLDGTLLFLSSVVRGS
VCLTAHIGSGVKAYDHPQGVISAFANEHGVALLRESATREPELVNGORVTHLHA
RFPFVREQRVTFEELGEGEWLLPGEOKVPALLNHGGPHDTYGGHFTHEFQL
MAARGVCYSNRPVSGVYGQAWDAIYGRKWTPDADLLNDFDRCLEAVPRDLAAKT
AVMGSGYGFMTNWTHTTRFQAATDRCISNLISFGGTSIDGLREWDDGLGDFSR
RADALKLDLSPLQYVENVKTPTLIHVSLDHRCPVEQAQVYAAALHKHGVPRVFRF
PEENHELSSRGRPDRLRLTRLNEYFAWLWL"
3098..3352
/gene="DR0166"
3098..3352
/gene="DR0166"
/note="similar to SP:P31787 PID:1323021 PID:1620454
PID:170970 percent identity: 66.09; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="acyl-CoA-binding protein"
/protein_id="AAF09755.1"
/db_xref="GI:6457835"
/translation="MNFQAQEQEVNTLSKKPNDVLLKLYALYKQGSAGVDGKRPPGG
PDFVGGAKYDAWAGLKGKTQEQAQVEYVALVELKAKDQ"
3663..4649
/gene="DR0167"
3663..4649
/gene="DR0167"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF09762.1"
/db_xref="GI:6457842"
/translation="MPSANVSPPCSGVRRGGMPKAKAEASKPHQIPVKLPFVTAP
DALAAKARMRLDAAAYVAALPGRDTHSLMAGVPGVDLAFMPLGRDGFDPHNVL
INSAAPERQRTLAHEHTGAILIGDDDLSDIHDAVEGERLOVLETLCNVAARAIL
MPEPVIAMLERFGTGRALAEALAKRAEVSASSALYALTETQTPVPIYAYCACPKPPR
EQASDEDAGPSTEKVLTIRASSSTGRVATLASGTFVPADHPAALALATAGMEVRES
YVPRSGRMKMAEDAYPSRGIVAVSFEPDPAARLGRKDEQADRDPEPQAAQ"
4646..5515
/gene="DR0168"
4646..5515
/gene="DR0168"
/note="similar to GB:L06494 SP:P26282 GB:U01376 GB:X68776
GB:X68777 percent identity: 62.16; identified by sequence
similarity; putative"
```

gene	gene
CDS	CDS
Query Match 9.7%; Score 191.2; DB 1; Length 10256; Best Local Similarity 55.7%; Pred. No. 1.le-49; Matches 389; Conservative 0; Mismatches 303; Indels 6; Gaps 1;	Query Match 9.7%; Score 191.2; DB 1; Length 10256; Best Local Similarity 55.7%; Pred. No. 1.le-49; Matches 389; Conservative 0; Mismatches 303; Indels 6; Gaps 1;
1272 GGAAGGTGAGACAAACATATCCACTTATCTTACATACACGGGGTCCGGATGATGTA 1331	1272 GGAAGGTGAGACAAACATATCCACTTATCTTACATACACGGGGTCCGGATGATGTA 1331
1787 GGAAGGTGAGCAGAAAGTGCCCGCTGCTCAATATTACGGGGGGCCACACCGACTA 1728	1787 GGAAGGTGAGCAGAAAGTGCCCGCTGCTCAATATTACGGGGGGCCACACCGACTA 1728
1332 CGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGCGAAAGGATACGGGTCGTTA 1391	1332 CGGACATACATATTTTCATGAGTTTCAGGTGCTGGCGGCGAAAGGATACGGGTCGTTA 1391
1727 CGGGCAGCGCTTACCCACAGATTCCAGCTGATGGCGGCGCGGTACCGGCTGCTGCTA 1668	1727 CGGGCAGCGCTTACCCACAGATTCCAGCTGATGGCGGCGCGGTACCGGCTGCTGCTA 1668
1392 TATCAATCCGAGAGGAAGCCACGGCTACGGGCGAGGAATTTGTGAATCGGTCAGAGGA 1451	1392 TATCAATCCGAGAGGAAGCCACGGCTACGGGCGAGGAATTTGTGAATCGGTCAGAGGA 1451
1667 CTCACCCCGCGGCTCGGTGGCTACGGCAGCGCTGGTGGACCCATCTACGGCG 1608	1667 CTCACCCCGCGGCTCGGTGGCTACGGCAGCGCTGGTGGACCCATCTACGGCG 1608
1452 TTATGGGGAAAGATTTATGACGATGATGACGCTGTCAGGCTGTGATGAGGCTATCAACGAGA 1511	1452 TTATGGGGAAAGATTTATGACGATGATGACGCTGTCAGGCTGTGATGAGGCTATCAACGAGA 1511
1607 CTGGGACGCTGACGCGGACGACGCTGCTGCACTTCTGACCGCTGCTGGAAGCGGT 1548	1607 CTGGGACGCTGACGCGGACGACGCTGCTGCACTTCTGACCGCTGCTGGAAGCGGT 1548
1512 TCCGATATTTGATCTTAAGCGGCTGCTGCTGACGCGGCGAGCTACGAGGCTTTATGAC 1571	1512 TCCGATATTTGATCTTAAGCGGCTGCTGCTGACGCGGCGAGCTACGAGGCTTTATGAC 1571
1547 GCGCGGCTGGACGCGCGCAAAACCGCGCTCATGGCGGCGAGCTACGCGGCTTCATGAC 1488	1547 GCGCGGCTGGACGCGCGCAAAACCGCGCTCATGGCGGCGAGCTACGCGGCTTCATGAC 1488
1572 CAACTGGATCGTGGGCGAGACGACCGCTTTAAAGCTGCCGTTACCCAGCGCTCGATATC 1631	1572 CAACTGGATCGTGGGCGAGACGACCGCTTTAAAGCTGCCGTTACCCAGCGCTCGATATC 1631
1487 CACTGATCACCAGGCGACACCGCTTTTCAGCGCGCATCACCGCGCTGCATCAG 1428	1487 CACTGATCACCAGGCGACACCGCTTTTCAGCGCGCATCACCGCGCTGCATCAG 1428
1632 AATTGATCAGCTTTTACGGGCTGATGATATCGGCTATTTCTTTACAGACTGGCAGCT 1691	1632 AATTGATCAGCTTTTACGGGCTGATGATATCGGCTATTTCTTTACAGACTGGCAGCT 1691
1427 CAACCTGATTTGTTTCGGGGGACGCTCCGACATCGGCTGCGCTTCTGGGACGACGAGCT 1368	1427 CAACCTGATTTGTTTCGGGGGACGCTCCGACATCGGCTGCGCTTCTGGGACGACGAGCT 1368
1692 TGACATCATCATGT-----TTGAGGACACAGAAAGCTCTGGACCGGCTCTCTTTAAA 1745	1692 TGACATCATCATGT-----TTGAGGACACAGAAAGCTCTGGACCGGCTCTCTTTAAA 1745
1367 GGGCTGGACTTCTCGCGCGGCGGACCGGCTGAAGCTCTGGGACCTTTCCGCACTGCA 1308	1367 GGGCTGGACTTCTCGCGCGGCGGACCGGCTGAAGCTCTGGGACCTTTCCGCACTGCA 1308
1746 ATACGAGCAAAACGTGGAGACACCGCTTTTGTATCTGATCGGCGGAGCGGATGACCGATG 1805	1746 ATACGAGCAAAACGTGGAGACACCGCTTTTGTATCTGATCGGCGGAGCGGATGACCGATG 1805
1307 ATAGTGGAAACGTGAAGCGCGCCCTGATCGTCCATTCGCTTGCACCGCGCTG 1248	1307 ATAGTGGAAACGTGAAGCGCGCCCTGATCGTCCATTCGCTTGCACCGCGCTG 1248
1806 CCCGATCAGAGCGGCGAGCAGCTGTTTATCGCTCTGAAATAATGGCAAGGAACCA 1865	1806 CCCGATCAGAGCGGCGAGCAGCTGTTTATCGCTCTGAAATAATGGCAAGGAACCA 1865
1247 CCCTGTGCAAGCAACCGAGCAGTGTGACGCGCGCTGCAAGCATCAGGTCGCGGTGCG 1188	1247 CCCTGTGCAAGCAACCGAGCAGTGTGACGCGCGCTGCAAGCATCAGGTCGCGGTGCG 1188
1866 GCTGTGCGTTTTCGGAATGATCGACAAATTTATCAGCAGCGGACACCAAGACAGCG 1925	1866 GCTGTGCGTTTTCGGAATGATCGACAAATTTATCAGCAGCGGACACCAAGACAGCG 1925
1187 CTTGCTGCGCTTCCCGGAGGAAACACGAACTGTCGGCTGCGCTGCGTGAAC 1090	1187 CTTGCTGCGCTTCCCGGAGGAAACACGAACTGTCGGCTGCGCTGCGTGAAC 1090
1926 GATCAAGCGCTGAATATATATCATGCTCATGCTTTGATC 1963	1926 GATCAAGCGCTGAATATATATATCATGCTCATGCTTTGATC 1963
1127 GCTACGGGCTGAACGAGTACTTTCGCTGCGTGAAC 1090	1127 GCTACGGGCTGAACGAGTACTTTCGCTGCGTGAAC 1090
RESULT 12	RESULT 12
AP000062	AP000062
LOCUS	LOCUS
DEFINITION	DEFINITION
ACCESSION	ACCESSION
VERSION	VERSION
KEYWORDS	KEYWORDS

SOURCE
Aeropyrum pernix
ORGANISM
Archaea: Crenarchaeota; Thermoprotei; Desulfurococcales;
Desulfurococcaeae; Aeropyrum.
REFERENCE
AUTHORS
1 (sites)
Kawarabayashi, Y., Hino, Y., Horikawa, H., Yamazaki, S., Haikawa, Y.,
Jin-no, K., Takahashi, M., Sekine, M., Baba, S., Ankai, A., Kosugi, H.,
Hosoyama, A., Fukui, S., Nagai, Y., Nishijima, K., Nakazawa, H.,
Takamiya, M., Masuda, S., Funahashi, T., Tanaka, T., Kudoh, Y.,
Yamazaki, J., Kushida, N., Oguchi, A., Aoki, K., Kubota, K.,
Nakamura, Y., Nomura, N., Sako, Y. and Kikuchi, H.
TITLE
Complete genome sequence of an aerobic hyper-thermophilic
crenarchaeon, Aeropyrum pernix K1
JOURNAL
DNA Res. 6 (2), 83-101 (1999)
MEDLINE
99310339
REFERENCE
2 (bases 1 to 251700)
Tanaka, T., Hino, Y., Kawarabayashi, Y. and Kikuchi, H.
AUTHORS
Direct Submission
TITLE
Submitted (14-DEC-1998) Yutaka Kawarabayashi, National Institute of
JOURNAL
Technology and Evaluation, Biotechnology Center; Nishihara 2-49-10,
Shibuya-ku, Tokyo 151-0066, Japan (E-mail: kyutaka@kazusa.or.jp,
Tel.: 81-3-3481-8951, Fax: 81-3-3481-8424)
COMMENT
Kawarabayashi, Y. is officially affiliated with the National
Institute of Bioscience and Human-Technology, Tsukuba, Ibaraki
305-0046, Japan.
Nomura, N. and Sako, Y. are at Kyoto University, Sakyo-ku, Kyoto
606-8502, Japan
The other authors are at the National Institute of Technology and
Evaluation, 2-49-10 Nishihara, Shibuya, Tokyo 151-0066, Japan.
All the sequence with length 100 codons or more between ATG or GTG
and stop codon are defined as CDS
Homology analysis is performed by Smith-Waterman algorithm against
GenBank and GenPept release 109; EMBL release 56.0; Swissprot
release 36.0; PIR-protein release 57.0; and OWL release 31.0.
E-mail address for comments and questions: genome@nite.go.jp
Restriction map, ORF organization, sequence alignment and more
information are available at W.W.W. site of Biotechnology Center,
URL: <http://www.miln.nite.go.jp/>.
Location/Qualifiers
FEATURES
source
1..251700
/organism="Aeropyrum pernix"
/strain="K1"
/db_xref="taxon:56636"
complement(948..1026)
/gene="APETRNA28"
complement(948..1026)
/gene="APETRNA28"
/product="tRNA-Gly"
/notes="codon recognized: GGG; anticodon: CCC"
/anticodon="(pos:988..990,aa:Gly)
1157..2149
/gene="APE1559"
1157..2149
/gene="APE1559"
/notes="similar to PIR:A69437 percent identity:31.228 in
285aa."
/codon_start=1
/transl_table=11
/product="330aa long hypothetical protein"
/protein_id="BAA80558.1"
/db_xref="GI:5105245"
/translation="MSGPGTYSPPSRLLQYLCCAGVAGARLVPGLGGRFFSNCLV
VEDAGFVLVDTCGPWVERGLVDVVDVVFTHPHDIRGFHRLRAGRVAAPGEPPYA
GRLERLAFGRERWISMARLVMDEGVPEGPYPRGEDLCFRGVCIKTYIPARG
HLQHTLLPGGLVHLVDIDLTGFGPWYGNPEASPELFLDIEAAEAREGYLSAH
KDRVSRGEGVEALAFAGRLVEOAGYVKALEAGRLWPWELTGRGIIYRRYLEGM
EALMSYFPGVMIEKLLGLLGLWGCTRKTPGGIVARDCHLGVIKDKITRNLISA"
complement(1196..1684)
/gene="APE1560"
complement(1196..1684)
/gene="APE1560"
/codon_start=1
/transl_table=11
gene
CDS
/product="162aa long hypothetical protein"
/protein_id="BAA80559.1"
/db_xref="GI:5105246"
/translation="MDKPPGQLEQGVLEPPSGVGLNADPPEAEVLPRPEILRLRH
AFQIHNPASHADPLPPPLAGKPEKPLKPSIGLLTLRGSHSPASQPMPDPWVVEV
SEHHYINHQPTLPLGLTPRIHENLEPRILNQHTVAKGKPPPKRPNKPSISHPSTAPQ
IL"
2258..2437
/gene="APES056"
2258..2437
/gene="APES056"
/notes="similar to OWL:SOU61050 percent identity:41.667 in
24aa."
/codon_start=1
/transl_table=11
/product="59aa long hypothetical protein"
/protein_id="BAA80560.1"
/db_xref="GI:5105247"
/translation="MARVRIVAKNGPILLEVKGVTALCRGHSNNKPKYCDGTGAR
VGRQADEKVAFAEF"
2536..2952
/gene="APE1561"
2536..2952
/gene="APE1561"
/notes="motif=neutral zinc metalloproteinases, zinc-binding
region signature"
/codon_start=1
/transl_table=11
/product="138aa long hypothetical protein"
/protein_id="BAA80561.1"
/db_xref="GI:5105248"
/translation="MRVERAPDVCRAAEVLVDALGDIDYRVYCVRSWGSGRAYA
RYGTFCGAALALGNGPIVIEVSEKFDPLPEEKVKVVIHELLHPIRTFTSGGLRAH
GPLVNGRRVAALYRRVGARVMGILREALDGGPQE"
complement(2599..3063)
/gene="APE1562"
complement(2599..3063)
/gene="APE1562"
/codon_start=1
/transl_table=11
/product="154aa long hypothetical protein"
/protein_id="BAA80562.1"
/db_xref="GI:5105249"
/translation="MGSRVSWPGRARPCYAGLGSWPFSPWPPPCWELFLASRI
OGFAEDAHNPGITPDPPVORCNPPVNCQAMGPEAAEGAGYVEQLMYDDNLNLLQWQ
RVELLEDNLVYAGVIAESRSPRPVDRISPTTPPPAPHAVHFGDVIDV"
complement(2906..3376)
/gene="APE1563"
complement(2906..3376)
/gene="APE1563"
/codon_start=1
/transl_table=11
/product="156aa long hypothetical protein"
/protein_id="BAA80563.1"
/db_xref="GI:5105250"
/translation="MEKEDGVLRLTTPPKGCDLTPDVHAYLGLSGVREIALGYTRV
YCRDGAEEVLDDLEGMLTAPGPGPIMATVTLSPGNAILLITLQERODGATAASTA
KGDAGEPGVAVREGREAGLRWGLGKLAILAMAGAAALLGAILLEGLHPRLRGGCP"
complement(3615..4268)
/gene="APE1564"
complement(3615..4268)
/gene="APE1564"
/notes="similar to PIR:H70375 percent identity:45.455 in
110aa, motif=C-5 cytosine-specific DNA methylases
signatures"
/codon_start=1
/transl_table=11
/product="217aa long hypothetical protein"
/protein_id="BAA80564.1"
/db_xref="GI:5105251"
/translation="MAGVGRPPPPSVGSVVVADTMTDAKLREVLGARSLEAARIM
ARILASRMGLTVITTSVRAEIERFMIGNGVHPTTIARLMAWVRVKPTTHLRLP
AMFRRYVETVARLDKGLURVALRDRDEGDIVRSRLREKRYETTRKGMDS

gene	VEDVDTLLALLEGAAIVTSDGLRRAEDLGLTLVLPVELLEY ILALEEVKSVEE"									
	3888..4304									
	/gene="APE1565"									
	3888..4304									
	/gene="APE1565"									
CDS	/codon_start=1									
	/transl_table=11									
	/product="138aa long hypothetical protein"									
	/protein_id="BA80565.1"									
	/db_xref="GI:5105252"									
gene	GNGCRHVAVQHEPLYLSPDARSVHRQPHSGSPGOYSPHYPGSLQAPGPDLAQLG									
	VKGHGRVCHHNAHARRRPPAHASHPLQPPCPSLLV"									
	3955..4299									
	/gene="APE1566"									
	/db_xref="GI:5105253"									
CDS	/product="114aa long hypothetical protein"									
	/protein_id="BA80566.1"									
	/db_xref="GI:5105253"									
	/translation="MAAGRRSCVVGGLTLTHAISLAMVVGCTPPSPMNLISALTIG									
	VMYTVSPILRARASILAIIAASSRLAPRTSRSLASVNMVVSATITLTLGGGLPPT									
gene	PATPSNPHPHF"									
	complement(4296..5405)									
	/gene="APE1567"									
	complement(4296..5405)									
	/gene="APE1567"									
CDS	/note="similar to Swiss_Prot:Q57857 percent									
	identity:33.430 in 344aa. motif=aminoacyl-transfer RNA									
	synthetases class-II signatures"									
	/codon_start=1									
	/transl_table=11									
	/product="369aa long hypothetical protein"									
	/protein_id="BA80567.1"									
	/db_xref="GI:5105254"									
	/translation="MDPGVDLEALSTRSVRSRFDVYVYVGRPRQVPEGLTVA									
	FRGEQIVVMGYPSIKMLLPVAPVPPFGPTVLVEKMGYNVRYTVLGMVYAT									
	RGGLKYTRRLRGLGALKILEDAEGSFAGEVVGLENYTRYTYVEAPGFG									
	YFIFDIKGGRLPPRYKFSLAPEYGLKTVNLLAEIPATASGVVERLYTIVEDLEKRG									
	EGVIVKDPGRVPELYKTSRTNIGDIRLGMVPPPEGRSFLPRLIRIFREWEIETR									
	RYGELGEAILAPATEAVEAVSRGRLVEFEELVANEVEAEVYAYFASLGVHLEIA									
	GVARGVDGVRAVPRKPKSEGEIARILETGISPLD"									
gene	complement(5525..5854)									
	/gene="APE1568"									
	complement(5525..5854)									
Query Match										
Best Local Similarity 8.2%; Score 161.8; DB 1; Length 251700;										
Matches 369; Conservative 0; Mismatches 302; Indels 6; Gaps 2;										
QY	1297	ATTCTTAA	CATACACGGCGGTCCGCATATGATGACGACATACATATTTTCATGAGTTT	1356						
Db	172547	ATTCTCTATATACATGGGGGCTTAAGACGAGCTACGCTACGCTTCATACACAGTTTC	172606							
QY	1357	CAGGTGCTGGCGGGAAGGATACGGCGTCTGTTTATATCAATCCGAGAGGAAGCCACGGC	1416							
Db	172607	CAAATGCTACGCTCTCAGGGCTTTGCGGTGATTTACTCGAACCCAGGGGAGTGACGGT	172666							
QY	1417	TACGGGCAGCAATTTGTGAATCGGTCACAGGAGATTATGGGGGAAGGATTATGACCAT	1476							
Db	172667	TACAGCGAGGTTGCGCCGAT---CTAAGGGGTAGATACGGTGTGACGATTACTCTGAG	172723							
QY	1477	GTGATCAGGCTGTGATGATGAGGCTATCAACAGGATCCGCATATTGATCCTCAAGCGGCTC	1536							
Db	172724	CTTATGAAGGTTGTTGACGAGGCCCTCTCGGACTTCCACACCTTGACCCGACAGGGGT	172783							
QY	1537	GGTGTACGGGCGGAAGCTACGAGGTTTTATGACCAACTGGATCGTCGGGACAGCAAC	1596							
Db	172784	GGTGTAAACGGGGGAGCTACGGGGGCTACATGACCAACGTTTATAGTAACCAAACTAGA	172843							

QY	1597	CGCTTTAAAGCTGCCGTTTACCCAGCGCTCGATATCAATTTGGATCAGCTTTCACGGCGTC	1656
Db	172844	AGGTTCAAGCGCGCGGTCACACAGAGGAGTGCAGCAACTGGATAAGCTTCTATGGCGAG	172903
QY	1657	AGTGATATCG ---GCTATTTCCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGGAGAC	1713
Db	172904	AGCGACATAGGATGGTATTTTCGCCCGGAACTTATCTCCGCGCAAGAACCATTGGAGAGAC	172963
QY	1714	ACAGAAAGCTCTGGGACCGGCTCTCTTTAAAATACGACGAAACGCTGGAGACACCGCTT	1773
Db	172964	CTCGAGAAATPACGTTTGACTTTAGCCCGCTCTCTCTCTGTGAGAATATTGAGACACCACTG	173023
QY	1774	TTGATCTGATCGGCGGAGGATGACCGATGCCGATGCCGATCGAGCGGAGCAGCTGTTT	1833
Db	173024	CTTATATATCATCCACAGAGGACTTTTAGATGCCCTCAGCGGTGATTTCAACTCTTC	173083
QY	1834	ATCGCTCTGAAAAAATGGCAAGAAACCAAGCTTGTCTCGCTTTTCCGAATGTCATCGCAC	1893
Db	173084	ACCGCTCTAAGCTGAAGGCTGTGGACACGAGCTTCTAGTGTTCAGGTGAGAACCAT	173143
QY	1894	AATTTATCAGCGACCGGACACCCAAAGACACGCGATCAAGCGCTTGAATATATCAGTCA	1953
Db	173144	AACCTCTCCGGAAGCGGAACCTCTAAGAGCGGCTGGCAAGGCTTAGGGCGATAGCCTCC	173203
QY	1954	TGCTTTGATCAACATCT 1970	
Db	173204	TGCTTCAAGGACACCT 173220	
RESULT 13			
AE010617/c			
LOCUS			
DEFINITION			
Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 159 of			
197 of the complete genome.			
ACCESSION			
AE010617 AE009951			
VERSION			
AE010617.1 GI:19714722			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
1 (bases 1 to 9531)			
AUTHORS			
Kapatral,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,			
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,			
Zhu,L., Vasileva,O., Chu,L., Kogan,Y., Chaga,O., Goltsman,E.,			
Bernal,A., Larsen,N., D'Souza,M., Walunas,T., Pusch,G.,			
Haselkorn,R., Fonstein,M., Kyrpides,N. and Overbeek,R.			
Genome sequence and analysis of the oral bacterium Fusobacterium			
nucleatum strain ATCC 25586			
TITLE			
J. Bacteriol. 184 (7), 2005-2018 (2002)			
JOURNAL			
MEDLINE			
PUBMED			
1189109			
REFERENCE			
2 (bases 1 to 9531)			
AUTHORS			
Kapatral,V., Anderson,I., Ivanova,N., Reznik,G., Los,T.,			
Lykidis,A., Bhattacharyya,A., Bartman,A., Gardner,W., Grechkin,G.,			
Zhu,L., Chu,L., Kogan,Y., Chaga,O., Goltsman,E., Bernal,A.,			
Larsen,N., D'Souza,M., Walunas,T., Pusch,G.D., Haselkorn,R.,			
Fonstein,M., Kyrpides,N. and Overbeek,R.			
Direct Submission			
TITLE			
Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park			
JOURNAL			
Drive, Chicago, IL 60612, USA			
FEATURES			
source			
1..9531			
/organism="Fusobacterium nucleatum subsp. nucleatum ATCC			
25586"			
/strain="ATCC 25586"			
/sub_species="nucleatum"			
/db_xref="ATCC:25586"			
/db_xref="taxon:190304"			
complement(114..2600)			
/gene="FN1122"			
complement(114..2600)			
/gene="FN1122"			


```

FTSHFGDYSIYLDPMANSLSTQYIHAFTRTTCGKQEIOWPSPISVCTPSIYG"
5619. 6149
/gene="PAE1000"
CDS
5619. 6149
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="P. aerophilum family 3 protein"
/protein_id="AAL63189.1"
/db_xref="GI:18159789"
/translation="MARLFKYPFRMGALGVMSFFHLFLFGLREGLLAAVSYFS
AYGVNWRVLPWGVAVGVFISGRIVVVFVFLYLAFLMVLPLFTRGVRSRVS
LFEVLGHFLSGVPLLLISPLWRSVPHPEFIAYGPAITPVDWPMWTFVGVYVWER
HSIKGLPSGFTREG"
6290. 6733
/gene="PAE1001"
CDS
6290. 6733
/note="Hypothetical; Conserved within genome"
/codon_start=1
/transl_table=11
/product="conserved within P. aerophilum"
/protein_id="AAL63190.1"
/db_xref="GI:18159790"
/translation="MRGAVVSIQVAVLLVFLIAVALAGLYLTYFYSLOYTYIALT
OAYTPPSGAEVKICFSGVSGGLTITGVELDRPAAEFKVIYRGAEAGEVRGDVG
YIRADFGVVRPQWMTGRVTRQGFSLTFPSVQAEGRCPE"
6760. 8643
/gene="PAE1002"
CDS
6760. 8643
/note="Protein fate; Protein modification and repair"
/codon_start=1
/transl_table=11
/product="acylamino-acid-releasing enzyme, conjectural"
/protein_id="AAL63191.1"
/db_xref="GI:18159791"
/translation="MSLSPELAKLVLSNPDAAGKGVFFTVTRISLENDREYSIWA
CEGVCKVPLPGPDTAARASPDGSKIAFLSRGDKDGVGLWVAWEGPEPROLAK
FLGVLDSPSGGALAVAYEGSPGDKVVERLPLINDFGFAYNYSHLYLVDSY
SGVKEITDGVDRVLAAMPDRRIATAVARDLLKYLQDVLVDLTKTGEMTAAAG
FSSVSDIAMSPPGHIAFTHLRPLRGLSSHRSIWELEGEPKRCVTCGFVNVGNSV
SDVSPGFARALYAGDGLLFQATVGTGVFRALSDGVDFVLPFGVDFEFLPG
AIVTYMEANPKELYIWDGEPRLTRFNDVFLERKLRRPSRLVLRASDGAEEVGG
VLLPFGPGKHWLYIHGGPKTAYGEGMFEHLIASRGYAVFNPGRSGYSEFA
DIRCKYGERDQDLMEVVDVLYKSPYLPNRCVAGVSGYGFMTWIIAHVDRFAAV
TQRCIDVMSYGTIDIGWYFVEDQLCCTPWRNRELCKEKSPLILANRYKTPILLHS
LEDYRTWVQGVLFFTALRLHGVETRLVLPFESHELTRKGRPRHVENFRKLLNLD
KYV"
8636. 9889
/gene="PAE1003"
CDS
8636. 9889
/note="Amino acid biosynthesis; Aspartate family"
/codon_start=1
/transl_table=11
/product="asparaginase (asnA)"
/protein_id="AAL63192.1"
/db_xref="GI:18159792"
/translation="MKIKRVVLENGDVEFGVMLPPTQFSDSDIVVLKKNYVGFK
KGRKEIVELGEVQTAPLAKPMKPIEGEKVWLLATGGTILSRVDYVTGGVYPTLSVD
YLFEVLGLEAPIAEAEVTAKEFSEDMTPALMGVIAERVAEAFKKGARGVVLHGDTM
QYTAALAFAPKSPGPIALVGAQRSSDPDVLNLKAAIAVTAAPFAESVVMH
KTSQDVTVAHRTGVRKMHTRSDTFOSINTPIAEYYPEKELLQVLDVYKERGGL
DYAKFEAEVALVFKYPCMHHPRLLEALLVGMKGVVIEGTGFHGVGGLPAVKKLID
AVRVLVTSQILYGRVNLVYRRGRELISMGVIPLEMLPETAYAKMSWALANFREER
VPRVLITPIAYEMSPRSDPLVFGL"
9886. 10416
/gene="PAE1004"
CDS
9886. 10416
/note="Hypothetical; Conserved"

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL63193.1"
/db_xref="GI:18159793"
/translation="MNDVIAVADTLCKTPRVGLQGVSEAESVCAHVLLTLLAGEI
AAHLNAGVEYDIKVAVAADHDLAESVLHGHPREVDRKIRWEAESEVFKREFPHL
AELFRWRYETNLVGRIVSFADKLATLRACRYKGLGDDDLIKLYRLAKYDDLA
HILDIVVARYCOGVTL"
BASE COUNT 2438 a 2847 c 3137 g 2811 t
ORIGIN
Query Match 8.0%; Score 158.2; DB 1; Length 11233;
Best Local Similarity 54.4%; Pred. No. 6e-39;
Matches 363; Conservative 0; Mismatches 298; Indels 6; Gaps 2;
QY 1304 ACATACACGGCGTCCGCATATGATGTACGGACATACATATTTTCATGAGTTTCAGGTGC 1363
|||||
DB 7979 ACATACACGGCGGGCCCCAAACGCTACGGCGAGGGGTATATTTTGAATTCACCTGC 8038
|||||
QY 1364 TGGCGCGGAAGGATACGGCGTCTTATATCAATCCGAGAGGAGCCACGGCTACGGGC 1423
|||||
DB 8039 TCGCGTCGGGGGATACGCTGTGTCTCAAAACCCGAGGGGAGCGGTATATCGG 8098
|||||
QY 1424 AGGAATTTGGAATCGGTCAGAGGAGATATGGGGGAAGGATATGACGATGTATGC 1483
|||||
DB 8099 AGGAGTTTGGCGAT---ATTAGTGTAAATACGGCGAGAGGGACTTTCAAGACTTAATGC 8155
|||||
QY 1484 AGGCTGTGATGAGGCTATCAACGAGATCCGCAATTTGATTCCTAAAGCGGCTCGGTGTCA 1543
|||||
DB 8156 AGTTGTGGACTACGTGCTTAAGAGCTATCGGC---TTGACCCCAATCGAGCGCTGTGG 8212
|||||
QY 1544 CGGCGGGAAGCTACGGAGGTTTATGACCAACTGGATCGTGGCGGACAGAACCGCTTTA 1603
|||||
DB 8213 CCGCGGGTGTACGGCGGCTTTATGACGAATTTGATAATTCGCCACGCTGCACAGTTTA 8272
|||||
QY 1604 AGCTCGCTTACCGAGCGCTCGATATCAATTTGATGATGATGATGATGATGATGATGATGAT 1663
|||||
DB 8273 GAGCCGCGGTGACGCGCGCTCAATATGCGACTGGGTGTCTATGTACGGCACGACGGGATA 8332
|||||
QY 1664 TCGGCTATTCTTTACAGACTGGCAGCTTGAGCATGATGATGATGATGATGATGATGATGATGAT 1723
|||||
DB 8333 TAGGTTGGTATTTGTGAGGATCAGCTCTGTTGCACTCCCTGGAGGAACAGGGAGCTCT 8392
|||||
QY 1724 TCTGGGACGGTCTCTCTTTAAATACGAGCAACCTGGAGGACACCGCTTTTGTATCTATGC 1783
|||||
DB 8393 CGCTTGAGAGAGCCCGCTTTATTTAGCCATAGGGTAAGACGCCGACATTAATTAATCC 8452
|||||
QY 1784 ATGGCGAGCGGGATGACCGATGCCCGATCGAGAGCGGAGCAGCTGTTTATCGCTCTGA 1843
|||||
DB 8453 ACTCTCTAGAGGACTACAGGACGTGGGTAGACAGGGGGTGTGTTCTTACGGCATTTGA 8512
|||||
QY 1844 AAAAAATGGGCAAGGAAACCAAGCTTGTCCGTTTTTCCGAATGCAATCGCACAAATTTATCAC 1903
|||||
DB 8513 GGCTTACGGGTTGAGACGGGCTGCTCTATTTCCCGGAGGAGACCGACGCTGACAC 8572
|||||
QY 1904 GCACCGGACACCAAGACAGCGGATCAAGCCCTGAAATATATATGCTCATGCTGTTGATC 1963
|||||
DB 8573 GCAAGGAAAGCCGCCGCCCGGGTGGAAAACTTTAAGGAGATCTTAAGCTGGCTGGATA 8632
|||||
QY 1964 AACATCT 1970
DB 8633 AGTATGT 8639

```

RESULT 15
 CNPAX05 304290 bp DNA linear BCT 07-AUG-2000
 LOCUS Pyrococcus abyssi complete genome; segment 5/6.
 DEFINITION Pyrococcus abyssi complete genome.
 ACCESSION AJ248287.1 GI:5458657
 VERSION AJ248287.1 GI:5458657
 KEYWORDS complete genome.
 SOURCE Pyrococcus abyssi.

ORGANISM

Pyrococcus abyssi
Archaea: Euryarchaeota; Thermococci; Thermococcales;
Thermococcaceae; Pyrococcus.
1 (bases 1 to 304290)
Heilig, R.
Pyrococcus abyssi genome sequence: insights into archaeal
chromosome structure and evolution
Unpublished
2 (bases 1 to 304290)
Genoscope.
Direct Submission
Submitted (06-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Location/Qualifiers

FEATURES

source

1..304290
/organism="Pyrococcus abyssi"
/strain="Orsay"
/db_xref="taxon:29292"
complement(77..2398)
/note="PAB1560"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation regulatory
protein (hypf)"
/protein_id="CAB50145.1"
/db_xref="GI:5458658"
/db_xref="SPTREMBL:Q9UZB4"

CDS

/translation="MKAYRIHVQIOVAVGFRPVPVYRIAHNLRGVRLGDAGVEI
VVGREDIEGLRDIYRKRPPIAKIEKVEREPIQGFDFRYIEKSTSEKKGEGDSI
IPDIAECDELREDFPTNRYMYPVICTNGCPRETIIDLPYDRENTAMREFFPM
DFCKSEYEDLNRYHAEPVACVPGPSYRLTSDGKEITGDLRKAALDKGYIIA
IKGIGGTALCATREDVVAELAKRLRFPCKPRAIMAKDLTVKGFAYISREBEELT
SYRPIVALKRPEPLPENLAPCLHTIGVMLPYAGTHYILFHWKTPVYVMTSANFP
GMPKIDNEAFKLDKVDVYLLHNRRIPNRADDSVVRVDGRRVIRSRGFVPLW
IEIPFETKGAELMNAFGVAKNGKVPYISQYIGNTSKIEVLEFMEAKIKHFKILR
INALDLVDLHPNNTYTKLAMEIAEFNAEFQVQHYYAHVASMAHLEEVGIA
LDVGYGDTGKTGGEIYLSYEDIERLAHLEYPLPGDLASYPLRALIGILSLNH
EIEVENTIRKFCPNALSKLYGETELKVLTRLOLSSGINVAYASSTGRVIDAELVLL
VSYRRYEGEPAMKLSFASQGNLDLKTTPVSGEIRVSELFEELDLMGKASPRI
AYSILHALAFAEVSVEKAKFGAKTVLGGVGVNELIVTKRIVEGNGLKFLPT
YEVRDNGINVGQAFGLGLLEGILNREDLST"
complement(2490..3593)
/note="PAB1559"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation protein (hypd)"
/protein_id="CAB50146.1"
/db_xref="GI:5458659"
/db_xref="SPTREMBL:Q9UZB3"

CDS

/translation="MLEKPRDKEIAQKIVRKIHKEAKGFDLRFMHVCGTHEDVTTRA
GIRSLLENVKIMSGPCPVCITPVEDIVRMMEIMRKAKEGEDLILTTGDMYRPT
PMGSFADLSEGDVRYVYSIDYSKIAQENRDKLVHFSFGFETITAPAGMLETVI
NEGDLNFKYSVHRLTPPAVEALLKAGTVPHGLIDPHVSTIIGVRGWAYLTETKYNIP
QVWAGPEPDLGLILIRMWKEGAKEINSEYRVVKNWGNVKAQELINKFFEVDKA
KWALGIIIPQSGLELKEWKLDEIRTLNPDVPLKLDLEKGLCGAILRGLALPPQCP
HFCKTTPRHPIGCMVSVGTCISIFKYGALF"
complement(3594..3821)
/note="PAB7315"
/codon_start=1
/transl_table=11
/product="hydrogenase expression/formation protein (hypc)"
/protein_id="CAB50147.1"
/db_xref="GI:5458660"
/db_xref="SPTREMBL:Q9UZB2"

CDS

/translation="MCLAIPIGRVVEIRDNVGVDFGVKREVRDLDSLVKVGVDYVIV
HTGFALKEDEERAKETLEAWDEVFSILGGE"
3949..4386
/note="PAB0821"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50148.1"

CDS

/db_xref="GI:5458661"
/db_xref="SPTREMBL:Q9UZB1"
/translation="MODSDSEIOELKRNLOVAFEVLAEIAQSVIRLNLVAYCGGI
GIDVAIPRNDPIAREIVKILFDLRRANVSQITRELKGRCKASRNTVRAKLNELKN
LGVVVEVPGERCKVYTLSDKDVYKRWLDLIGIPKLEHINDYLR"
4392..4856
/note="PAB0822"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50149.1"
/db_xref="GI:5458662"
/db_xref="SPTREMBL:Q9UZB0"
/translation="MEKEKLEKVEELIEIKSAKTEEEVEVLRRKAVKKLIEAYE
GDKDLTKIAMLMDKLGPMVEDILPELQKLADLYSPERMOALGKSAQFYKDLVEAGM
DKETATELTKFPMKNINVMRSLEGLTGMLOKGGQFYVYVHGKKEKRETKS"
complement(5039..5668)
/note="PAB1558"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50150.1"
/db_xref="GI:5458663"
/db_xref="SPTREMBL:Q9UZA9"
/translation="MLKEIFNGIRFGEIALLJEYSNTPTVYVHLHLLSWSEKGYKIV
IFDVFDSLIVYRRKAEELGFDTSICKAKVYKVGVRNMGVNMKLTVTYEGPFEKVL
FEATEOVYEDKTIIVPLGTKEVLSLYPPRELOSTNFIALRVGDKRKAFFHINTDLV
KAVAPQILPMLSESTFVMLRKLYGKVEKLIWAKSLNPKLDGLEVLAKS"
complement(5709..6227)
/note="PAB1557"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50151.1"
/db_xref="GI:5458664"
/db_xref="SPTREMBL:Q9UZA8"
/translation="MEGRLLYPMRTYLVVSGHDEANVMAADVTVLSTDFPFVGVGA
IAPRTHKLKKINEPVSIPSEMLRDWVIAGTKGPKSKDKMGITLVVPSRKVKVP
SIRKALANICEVYTDYDGHITLVFGVKLAYTYKEAEFEGGKPNLKFNLARHSWE
FVTFQDKIYRAE"
complement(6269..6913)
/note="PAB1556"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="CAB50152.1"
/db_xref="GI:5458665"
/db_xref="SPTREMBL:Q9UZA7"
/translation="MSVEFSTMILLIMVLAAVSTLQFYKGBKLMLLMQHYLSIEAV
VKPKDXYVMVGYYGFRAYKKLEDKDILFEYTLTLPLPROSILYFPISLLINRHKIL
YVVRPTDILKREVHFKQGYFRMKPRIERIELLKGEIEINGVKFELYKSRDVEA
VKRFISGFSKVKVKVHALVPKTNVIYVFFKPEPESEDVVRNIIRFVKRMANR"
complement(7184..8176)
/note="PAB1555"
/codon_start=1
/transl_table=11
/product="ANION TRANSPORTING ATPASE"
/protein_id="CAB50153.1"
/db_xref="GI:5458666"
/db_xref="SPTREMBL:Q9UZA6"
/translation="MREFLTPKEGFRVIFVIGKGVGKTTSSASISVALAKAGYRTLL
VSLDPAHNLGDVMEKLSDKPREILDNLVASELMEKIGYLEHLKKNLKNMRYTLT
VINLEKFEVLYRSPGLEIATLEAIRLVLGDWDVIVFDTPPTGLTLVLALPRI
SLWTEKLDIDIRKKILERRAITKIQGEOKFVIDGEIKVPTREEDAVMRELKAYRD
EIKFEVDLTPNPKTSIAYVMNPEFLPYETKRAYESLKKFKIPFNVMVNMKVLKLR
EIPERVKLEAQEKVSEVKRPFPGVEIPEIPQEEPRGIEWLSKVGMETVGR"
complement(8489..10267)
/note="PAB1554"
/codon_start=1
/transl_table=11
/product="CARBON STARVATION PROTEIN A."
/protein_id="CAB50154.1"
/db_xref="GI:5458667"

```
/db_xref="SPTREMBL:Q9UZA5"  
/translation="MNSTVILAAIIIVGLYFTYTGKLGKLVVADPNRRTPAHRLY  
DGYDVPFAHPLVLYGHFASITAGSRPIVGPVAVAMAWGLPSLIWVFNFGVIGAVHDY  
LALMSSVYDQKSIQWIAKLSMRTSMAFELYVMFALILVIAFAAVIAGVFKTPG  
AAATLIFLLIALLVLLWLLYKQINFRVGTIGLIIAIYIGFKFPLVASAKAWYI  
FLGAYIVASSLPVWLLQPRDYNALVILWFLGLGFLAFIIGFKGVGTFTAPAYTT  
FSRVVGGKSPSPWPTIPLVACAGLSGFSLVSGTSSKOLDNLHGLLVGYGMEFT  
EGSLTIVISIAVIGLQIFHNAGIGITAETWQAAYAPLVAAKSITTETLQKLGVAAL  
ENFKNGTVIKGVIFASSYAGLHDAFDGDIHLGLTFASLWVSAPAFITLDTATKLRG  
FAWQEIITEMVKWPEILKNKWASPIVFFGIALAKWKWLVIPAFSGMGNQWLASIAM  
MTASLWMLKQRPQSGFWKWAFLIPALFLWTVTSALIFLWLVVGVKSSIVWSLITIVGL  
VLELLVIDWYNAWKPKPAEEYQAAAA"  
11287..13068  
/note="PAB0824"  
/codon_start=1  
/transl_table=11  
/product="RNASE L INHIBITOR"  
/protein_id="CAB50155.1"  
/db_xref="GI:5458668"  
/db_xref="SPTREMBL:Q9UZA4"  
/translation="MYRMRKRIAVIDYKCNPDKCGHFLCERVCVNRMGGEALIIDEE  
NYKPIQEAACSTGGCIVHKCPFNAISIVNPEQDEDCVHRIGYNAFVLRLPIVKD  
GMVGIIVPGTGTATVAKILAGOLPNLCEDNWDNVIRAFRGNELONYFERLUNG  
EIRVAPQYVLLKPAVKVRELLKVEVEVKELEVENLDRELHQLSGG  
ELQVATAALLRAHFVDFEPSYLDIQRKLVARVIRLANEGKAVLVVYHQLAV  
LDYLSDVHYVYGPVYGVIFPKPTRNGINFLQYKLDENVFRFYETFTKLSE  
RDYVERETLYPRLVKDYGSFKLEVEPEIRKGEVIGIVGPNIGKTFVKMLAGVE  
EPTGRKVDLTVAYKQYIKAEYEGTVYELLSKLDSKLSNFYKTELLKPLGLIDL  
YDRNVEDLSGELQVATAATLLRDADYLLIDEPSAYLDVFORLAVSRAIRHLMKNE  
KTALVDEHDLIMDYSDRLVIFEGEPGRHGRALPPMGMRGEMNRFASVGTIFRRDP  
DSGRPRANKESVDRQEKARGEYVYA"  
13130..13513  
/note="PAB0825"  
/codon_start=1  
/transl_table=11  
/product="TRANSLATION INITIATION INHIBITOR, PUTATIVE."  
/protein_id="CAB50156.1"  
/db_xref="GI:5458669"  
/db_xref="SPTREMBL:Q9UZA3"  
/translation="MAKEVIEITEKAPKPIGYSQAIKVGNEIFVAGQIPIDPETGEIV  
KGDYKEQTKRVNIENIKALEEAGASLNDVVKVTVYKLDLNDFAKMNENYSEYFGSKP  
ARVAEVSRLPKDVLIEMEATAYKE"  
13623..14480  
/note="PAB0826"  
/codon_start=1  
/transl_table=11  
/product="hypothetical protein"  
/protein_id="CAB50157.1"
```

Query Match 8.0%; Score 158; DB 1; Length 304290;

Best Local Similarity 52.7%; Pred. No. 1.9e-38;

Matches 391; Conservative 0; Mismatches 345; Indels 6; Gaps 2;

```
QY 1230 CGCGGTGATGGTGAACGGCTGGCTGATGAGGCGCTGCACAAATGGAAGGTGACACACATA 1289  
DB 209349 CGAGTTGGAGATAGACGGCTGGTACATAAAGCCGAGCTCAAGGGGGGGGAGAGGC--- 209293  
QY 1290 TCCACTTATCTTAACATACACGGCGGCTCCGCTATATGATGACGGACATACATATTTCA 1349  
DB 209292 TCCGGTATATAGTCTTCGTCACGGTGGGCCCAAGGAATGTACGGTTACTACTTCAAGTA 209233  
QY 1350 TGAGTTTCAGGTGCTGGCGGCGAAAGGATACCGGGTCGTTTATATCAATCCGAGAGGAAG 1409  
DB 209232 CGAGATGCAGCTTATGGCGAGCAAGGGCTACTACATAGTCTATGTCAATCCCAAGGGGAAG 209173  
QY 1410 CCACGGCTACGGGAGGAAATTTGTGATCCGGTCAGAGGAGATTATGGGGGAAGGATTA 1469  
DB 209172 CAACGGTTACACGAGGAGCTTTGACCTTAGCTTACGAGAGGACTGGCCCTAGAGGACT 209113  
QY 1470 TCACGATGTGATGAGGCTGTGGATGAGGCTATCAACACGAGATCCGCATATTTGATCCTAA 1529  
DB 209112 CCAGATATACTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 209053
```

Search completed: July 4, 2003, 13:12:17

Job time : 5004 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:58:29 ; Search time 452 Seconds
(without alignments)
9820.108 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaaaaagctataaccgc.....catggttgatcaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SID22/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1971	100.0	1971	AA07301	Bacillus subtilis
2	807.8	41.0	1958	ABK72933	Bacillus lichenifo
3	238.2	12.1	265118	AAH41227	Pyrococcus abyssi
4	227.2	11.5	1896	AAV33582	Pyrococcus horikos
5	195.6	9.9	1032	ABQ86198	Lactobacillus rham
6	192.2	9.8	477	ABK77575	Bacillus clausii g
7	158	8.0	349980	AAH41226	Pyrococcus abyssi
8	149.8	7.6	1869	AAV12887	Thermococcus amida
9	82.4	4.2	585	ABK77547	Bacillus clausii g

10	49.8	2.5	659	21	AAF13721
11	46.4	2.4	1797	24	ABK73047
12	39	2.0	6741	21	AAA10595
c 13	38.6	2.0	4590	22	AAH24065
14	37.4	1.9	2196	12	AAQ10958
c 15	36.6	1.9	788	24	ABO29152
16	36.6	1.9	788	24	ABO29153
c 17	36.4	1.8	2365589	24	ABA90521
c 18	36.2	1.8	9271	19	AAV23080
19	36	1.8	568	24	ABK75665
20	35.8	1.8	1395	18	AAV7586
21	35.4	1.8	716	22	AAH08175
c 22	35.2	1.8	8846	22	ABAJ5292
23	34.8	1.8	2196	13	AAO20001
24	34.6	1.8	45613	22	AAF28535
c 25	34.2	1.7	729	20	AAZ28861
26	34.2	1.7	10732	21	AAA10594
27	34.2	1.7	15164	24	ABK35500
28	33.6	1.7	420	16	AAO88233
29	33.6	1.7	446	21	AAH30728
c 30	33.6	1.7	2301	23	ABL24372
c 31	33.6	1.7	5828	22	AAF89816
c 32	33.6	1.7	5828	24	ABK84150
c 33	33.6	1.7	5828	24	ABN96524
c 34	33.6	1.7	6151	16	AAO75977
c 35	33.6	1.7	6522	24	ABL33803
c 36	33.6	1.7	20300	24	ABK47337
c 37	33.4	1.7	744	21	AAZ53852
38	33.4	1.7	957	20	AAZ06344
39	33.4	1.7	2176	23	AAH89363
40	33.4	1.7	2277	19	AAV05370
41	33.4	1.7	3549	22	AAH45086
42	33.4	1.7	4590	22	AAH24065
c 43	33.4	1.7	4917	23	AAH89366
c 44	33.2	1.7	1406	23	ABL06013
c 45	33.2	1.7	3480	23	ABL06012

ALIGNMENTS

RESULT 1
AA07301
ID AA07301 standard; DNA; 1971 BP.
AC AA07301;
XX
XX 21-MAY-1999 (first entry)
DT
XX
XX
DE Bacillus subtilis serine protease SP1 (YuxL) DNA.
XX
KW Serine protease SP1; YuxL; detergent; ss.
XX
XX Bacillus subtilis.
XX
XX WO9903984-A2.
XX
XX 28-JAN-1999.
XX
XX 14-JUL-1998; 98WO-US14647.
XX
XX 15-JUL-1997; 97EP-0305232.
XX
XX (GENV) GENECOR INT BV.
XX (GENV) GENECOR INT INC.
XX Estell DA;
XX
XX WPI; 1999-132231/11.
XX P-PSDB; AAW97789.
XX
XX Use of serine protease genes from Gram-positive microorganisms - for
PT modification of host cells for the production of heterologous

PT proteins or for producing proteins for use in cleaning compositions

PS Disclosure; Fig 1A-C; 37pp; English.

XX This DNA sequence encodes serine protease SP1 (YuxL) of *Bacillus subtilis*. Novel serine proteases SP1, SP2, SP3, SP4 and SP5 (see AA057789-93) were identified via a FASTA search of *Bacillus subtilis* genomic nucleic acid sequences. SP1 was identified by its structural homology to the S9 type serine protease dap2 of yeast. SP2, SP3, SP4 and SP5 were identified by their structural and overall amino acid homology to SP1. Host cells in which the naturally occurring gene encoding one or more of SP1, SP2, SP3, SP4 or SP5 is mutated such that the proteolytic activity is diminished or deleted altogether, can be used for the production of heterologous proteins, e.g. a hormone, enzyme, growth factor, cytokine, protease, carboxylase, lipase, racemase, epimerase, tautomerase, mutase, transferase, kinase or phosphatase (claimed). SP1, SP2, SP3, SP4 and SP5 can also be produced on a large scale in a microbial host expression system for use in cleaning compositions such as detergents, bar or liquid soap, dish-care formulations and contact lens cleaning solutions, or for peptide hydrolysis, waste treatment, textile applications, as fusion-cleavage enzymes in protein production, and as animal feed additives.

XX Sequence 1971 BP; 558 A; 431 C; 524 G; 458 T; 0 other;

Query Match 100.0%; Score 1971; DB 20; Length 1971;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1971; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	ATGAAAAGCTGATAACCGCAGAGGACATCACAGGATGTCTCTGTGACCGATCCTCAA	60
DB	1	ATGAAAAGCTGATAACCGCAGAGGACATCACAGGATGTCTCTGTGACCGATCCTCAA	60
QY	61	TAGCGCCGACAGGTCACCGGCGGATATGTAATCAACAAGTAATCAAGAAAGAT	120
DB	61	TAGCGCCGACAGGTCACCGGCGGATATGTAATCAACAAGTAATCAAGAAAGAT	120
QY	121	TCGTATACATCAATATATGATCTATGAAACGAAACGGGAGGATCTGTTCTTGGACA	180
DB	121	TCGTATACATCAATATATGATCTATGAAACGAAACGGGAGGATCTGTTCTTGGACA	180
QY	181	CATGGAGAAAAGGAAAGCAGCCGACCAAGATGTTCTCCGGAGGCGGACGCTTGCCTTT	240
DB	181	CATGGAGAAAAGGAAAGCAGCCGACCAAGATGTTCTCCGGAGGCGGACGCTTGCCTTT	240
QY	241	ATTCTCATGAGAGCGGATCGGCACAGCTTTATATCATGACGACTGAAGCGGAGAA	300
DB	241	ATTCTCATGAGAGCGGATCGGCACAGCTTTATATCATGAGCAGCTGAAGCGGAGAA	300
QY	301	GCAAGAAACTGACTGATATCCATATGGCGTGTCAAGCGGCTATGTTCCCGGACGGT	360
DB	301	GCAAGAAACTGACTGATATCCATATGGCGTGTCAAGCGGCTATGTTCCCGGACGGT	360
QY	361	GAATCGATCTGTCATATCAGTTTGGAGAGGGGAAAGCATTGATGACCGGAGAAA	420
DB	361	GAATCGATCTGTCATATCAGTTTGGAGAGGGGAAAGCATTGATGACCGGAGAAA	420
QY	421	ACAGAGCAGACAGCTATGAACCTGTTGAAGTGCAAGGCTCTCTCAACACGGGACGGC	480
DB	421	ACAGAGCAGACAGCTATGAACCTGTTGAAGTGCAAGGCTCTCTCAACACGGGACGGC	480
QY	481	AAAGGCTGACGAGAGTGCGTATGCCAGCTTGCTGTGTCAGCGTAAAGTCGGGTGAG	540
DB	481	AAAGGCTGACGAGAGTGCGTATGCCAGCTTGCTGTGTCAGCGTAAAGTCGGGTGAG	540
QY	541	ATGAAAGAGCTGACAAAGTCAAAAGCTGATCATGGTGATCCTGTTTTCTCTGACGGC	600
DB	541	ATGAAAGAGCTGACAAAGTCAAAAGCTGATCATGGTGATCCTGTTTTCTCTGACGGC	600
QY	601	AAATGCTGTTTTCTCAGCTAATTTAACTGAACAGATGATGCCACCAAGCCGATGAT	660
DB			

DB	601	AAATGCTGTTTTCTCAGCTAATTTAACTGAACAGATGATGCCAAGCCGATGAT	660
QY	661	GTTTACATAATGTCACCTGGAGTCTTAAGCAGGTACACCTCATCGCGGTCA	720
DB	661	GTTTACATAATGTCACCTGGAGTCTTAAGCAGGTACACCTCATCGCGGTCA	720
QY	721	TTCCGATCAAGCTCAITTTTACCAGACGGAAGTATCTTCTTGTGTTGAAATGAAAG	780
DB	721	TTCCGATCAAGCTCAITTTTACCAGACGGAAGTATCTTCTTGTGTTGAAATGAAAG	780
QY	781	GAATATAAGAACTGCTACGCTCTCAAGGGCTGCTATGATATCGAACAGCCGCTC	840
DB	781	GAATATAAGAACTGCTACGCTCTCAAGGGCTGCTATGATATCGAACAGCCGCTC	840
QY	841	ACATGCTTACTGAGATGCTGGACGTTTATTAGCGGATCGCTGATTGAGATTCATTG	900
DB	841	ACATGCTTACTGAGATGCTGGACGTTTATTAGCGGATCGCTGATTGAGATTCATTG	900
QY	901	ATCGGTGCTGTGAACAGCCCGGATTTGGACAAAGGACAGCCAAAGGTTTTATGTCATC	960
DB	901	ATCGGTGCTGTGAACAGCCCGGATTTGGACAAAGGACAGCCAAAGGTTTTATGTCATC	960
QY	961	GGCAGAGATCAAGGCGAGTACGGGCATCTATTATTTTCGATTGAAGCCCTGTGATCCG	1020
DB	961	GGCAGAGATCAAGGCGAGTACGGGCATCTATTATTTTCGATTGAAGCCCTGTGATCCG	1020
QY	1021	ATTGCTCTGAAAAAGAGTACATCAATAGCTTTCTTTCACCTGATGAACAGCACTTT	1080
DB	1021	ATTGCTCTGAAAAAGAGTACATCAATAGCTTTCTTTCACCTGATGAACAGCACTTT	1080
QY	1081	ATTGCCAGTGTGACAAAGCCGGACAGCCAGTGTTCACAGTATCCCGCTTGGACAG	1140
DB	1081	ATTGCCAGTGTGACAAAGCCGGACAGCCAGTGTTCACAGTATCCCGCTTGGACAG	1140
QY	1141	GAAGAGAAACAGTCTGACTGGCGGATGACAAAGTGTTCAGGAGCATATCATATCAATA	1200
DB	1141	GAAGAGAAACAGTCTGACTGGCGGATGACAAAGTGTTCAGGAGCATATCATATCAATA	1200
QY	1201	CCTGAAGAGATTCATATGCTACAGAAAGCCGGTGTGTTGTAAGCTGGCTGATGAGG	1260
DB	1201	CCTGAAGAGATTCATATGCTACAGAAAGCCGGTGTGTTGTAAGCTGGCTGATGAGG	1260
QY	1261	CCTGCACAAATGGAAGTGAGACAATATTCACCTTATTTTAACTATACAGCGGCTCGG	1320
DB	1261	CCTGCACAAATGGAAGTGAGACAATATTCACCTTATTTTAACTATACAGCGGCTCGG	1320
QY	1321	CATATGATCTACGGACATACATATTTTCAGTGTTCAGTGTGCGCGGAAAGGATAC	1380
DB	1321	CATATGATCTACGGACATACATATTTTCAGTGTTCAGTGTGCGCGGAAAGGATAC	1380
QY	1381	GGGGTCTTTTATATCAATCCGAGAGGAGCCAGGCTACGGGAGGAAATTTGTAATGCG	1440
DB	1381	GGGGTCTTTTATATCAATCCGAGAGGAGCCAGGCTACGGGAGGAAATTTGTAATGCG	1440
QY	1441	GTCAGAGAGATTTATGGGGAAGGATATGACGATGTGATGACGCTGTGGATGAGGCT	1500
DB	1441	GTCAGAGAGATTTATGGGGAAGGATATGACGATGTGATGACGCTGTGGATGAGGCT	1500
QY	1501	ATCAACAGAGATCCGATATTTGATCTTAAGCGGCTGCTGTCAGCGGCAAGCTACGGA	1560
DB	1501	ATCAACAGAGATCCGATATTTGATCTTAAGCGGCTGCTGTCAGCGGCAAGCTACGGA	1560
QY	1561	GGTTTTATACCAACTGGATCTCGGGCAGACAGCCGCTTTTAAAGCTGCCGTTACCCAG	1620
DB	1561	GGTTTTATACCAACTGGATCTCGGGCAGACAGCCGCTTTTAAAGCTGCCGTTACCCAG	1620
QY	1621	CGCTCGATATCAAAATGGATCAGCTTTTCAGCGGCTCAGTGATATTCGCTTTTACA	1680
DB	1621	CGCTCGATATCAAAATGGATCAGCTTTTCAGCGGCTCAGTGATATTCGCTTTTACA	1680
QY	1681	GACTGGCAGCTTGGACATGACATGTTTGGAGGACAGAAAGCTCTGGGACCGCTTCCT	1740
DB	1681	GACTGGCAGCTTGGACATGACATGTTTGGAGGACAGAAAGCTCTGGGACCGCTTCCT	1740


```
Db 58671 TTAGCCTGGAGGAAGTATCTCTGCACTGCTGGAATTCAGGTGGCCAAACACACCT 58612
Qy 1331 ACGGACATACATATTTTCATGAGTTTCAGTGTCTGGCGGGAAGGATACGGGTCTGTTT 1390
Db 58611 ACGGTTATGCTATTTATGACGAGTTTCAGGTTTATGTTGCAAGGCTTCGTTGTAATCT 58552
Qy 1391 ATATCAATCCGAGGGAAGCCAGCGCTACGGGACGGAATTTGTCATGCGGTTCAGAGAG 1450
Db 58551 TTTCAATCCAGGGGAAGTGATGGTTATGGTGAGGAGTTTCGGG---GACATAGGGGTC 58495
Qy 1451 ATTATGGGGAAGGATTTATGACGATGTGATGACAGGCTGTGGATGAGGCTATCAACGAG 1510
Db 58494 ACTACGGTGAGCGGATTTATCAAGACCTAATGAGGTAGTTGACGAGGCTTTAAAGCCT 58435
Qy 1511 ATCCGCATATTGATCTTAAGCGGCTCGGTGTACGGGCGGAAGCTACGGAGGTTTATGA 1570
Db 58434 TTGACTTTATAGATGAGAAAGGCTTGCGTTTACCGGAGGATCATCGGAGGCTTCATGA 58375
Qy 1571 CCAACTGGATCGTCGGGACAGCAACCGCTTTTAAAGCTGCGGTACCCAGCGCTCGATAT 1630
Db 58374 CAAATGGATCGTTGGGACACAAACAGATTCAAAGCTGCGTAACCTACGCGCTCAATAT 58315
Qy 1631 CAAATGGATCAGCTTTACGGGCTCAGTGATATCGGCTATTTCTTTACAGACTGGCAGC 1690
Db 58314 CCAACTGGATCAGCTCTTTTGAAGACCGCATATCGGCTATTTACTTCGCTCCAGATCAGA 58255
Qy 1691 TTGAGCATGATGTTTGGAGACACAGAAAGCTCTGGGACCGGTCTCCTTTTAAATACG 1750
Db 58254 TCGGTGGGATCTCTTGGAGTAACCTAGAGGCTACTGGGAGAGAGCCCATTTGAAGTACG 58195
Qy 1751 CAGCAACGTTGGAGACACCGCTTTTGATCTGATGCGGAGCGGATGACCGTGCCTGA 1810
Db 58194 CTCCAACGTTGAGACTCCGCTCTGATATTTCACTCCAGGAGGACTACGCTGTGTCG 58135
Qy 1811 TCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATTTGGCAAGGAACCAAGCTTG 1870
Db 58134 TTCCGAGGCTTTGCAATTTTACATAGCGTTGAAGTATTTTGGGCAAAACCGTCGAGCTG 58075
Qy 1871 TCGGTTTCCGAATGCATCCACAAATTTATCAGCACCAGCGGACACCCGAAGCAGGATCA 1930
Db 58074 CGATATTCACAGGTGAGAAATCATGACTTGAGTAGGAGTGAAGCCCAAGCATAGGTTA 58015
Qy 1931 AGCGCTTGAATTTATATCAGCTCATGTTTGA 1961
Db 58014 AGAGCTCGAGCTCATAGTGGGTGTTGA 57984
```

RESULT 4

```
AAV33582
ID AAV33582 standard; DNA; 1896 BP.
XX
AC AAV33582;
XX
DT 29-DEC-1998 (first entry)
XX
DE Pyrococcus horikoshi acylpeptide hydrolase encoding DNA sequence.
XX
KW Pyrococcus horikoshi acylpeptide hydrolase; ss.
XX
OS Pyrococcus horikoshi.
XX
FH Key Location/Qualifiers
FT CDS 1..1896
FT /product= "Pyrococcus horikoshi acylpeptide hydrolase"
FT /note= "CDS does not contain a stop codon"
```

JPI0210977-A.

11-AUG-1998.

```
31-JAN-1997; 97JP-0018381.
```

```
XX 31-JAN-1997; 97JP-0018381.
XX (AGEN ) AGENCY OF IND SCI & TECHNOLOGY.
XX WPI: 1998-488369/42.
XX P-PSDB; AAW70508.
XX A heat-resistant acylpeptide hydrolase and a gene coding it - useful
XX for hydrolysing the C termini of proteins at high temperature
XX Example 5; Pages 6-7; 8pp; Japanese.
XX The present sequence represents a Pyrococcus horikoshi acylpeptide
XX hydrolase encoding DNA sequence. The enzyme has the following
XX properties: (a) it hydrolyses an acylpeptide; (b) an optimum temperature
XX of 90-95 degrees Centigrade; (c) an optimum pH of 5.0-6.0; (d) no loss
XX of activity when heated to 95 degrees Centigrade for 3 hours at pH 7.5;
XX and (e) a molecular weight of 60 kDa. The enzyme is claimed useful
XX for hydrolysing the amino end of an acylated protein and a peptide at
XX high temperatures.
XX Sequence 1896 BP; 614 A; 358 C; 491 G; 433 T; 0 other;
SQ
Query Match 11.5%; Score 227.2; DB 19; Length 1896;
Best Local Similarity 55.4%; Pred. No. 1.2e-61;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;
Qy 1139 AGGAAGAGAAACAGCTGACTGCGCGAATGACAAAGTTTGTTCAGGAGCATACGATATCAA 1198
Db 1061 ATGCAAGGAGGAAGGTTTACCACCTTTTAAACAAATGATAAAGGTTTACACCCCTTCAA 1120
Qy 1199 TACCTGAGAGATTCAATATGCTACAGAGAGCGCGTATGTTGTAACGCGGTGCTGATGA 1258
Db 1121 AACCTGAACACTTTTAAAGGTTTAAAGCAAGTACGCGGTTTGAATAGATGCTGGGTAATGA 1180
Qy 1259 GGCCTGCACAAATGGAAGGTGAGACACATATCCACTTTATCTTAAACATACACGCGGTC 1318
Db 1181 AACCGGTGAACCTTCAGGAAAGAAAGATATCCAGCTATTTCTAGAGATCCACGTTGTC 1240
Qy 1319 CGCATATGATGTACGGACATACATATTTTCATGATTTTCAGTGTCTGCGGCGGAAGAT 1378
Db 1241 CTAAACCGCTTACGGTTTACGCTTTTATGACAGATTCCACGCTTTTAAACCTCTAAAGCT 1300
Qy 1379 ACGCGTCTGTTTATATCAATCCGAGAGAACCCAGCGTACGGCAGCAATTTGTAATG 1438
Db 1301 TCGTCTGATATTTCAAACTCTAGAGGAGCGATGCTACGGAGAGGATTCGCGGAT- 1359
Qy 1439 CGGTACAGAGGATTTATGGGGAAGGATTTATGACGATGTGATGACGCTGTGGATGAGG 1498
Db 1360 --ATAAGGGGACACTATGGGAGAGGATTTACCAGGATTTAATGAGGTAGTCGATGAAG 1417
Qy 1499 CTATCAACAGAGATCCGCATATTTGATCTTAAGCGGCTCGGTGTCTACGGGCGGAAGCTACG 1558
Db 1418 CATTAAGGAGATTTGACTTTCATAGATGGGAAAGGCTAGGAGTTTACCGGGGTTCTCTATG 1477
Qy 1559 GAGGTTTATACCAACTGGATCTCGGCGACACCAACCGCTTTTAAAGCTCCGCTTACCC 1618
Db 1478 GTGCTTCATGACGAACCTGGATAGTCGGACATACCAACAGGTTTCAAGCCGCTGTAACCC 1537
Qy 1619 AGCGCTCGATATCAAAATTTGATCAGCTTTTCAAGCGGCTCAGTGATATCGGCTATTTCTTTA 1678
Db 1538 AGAGATCAATTTCAAAATTTGATAGCTTCTTCGGGACACACGATATAGTTTATTTTG 1597
Qy 1679 CAGACTGGCAGCTTGAGCATGACATGTTTGTAGGACACAGAAAGCTCTGGGACCGGCTCTC 1738
Db 1598 CTCAGATCAAAATAGGAAAGATCCCTGGAGCAACTTGGAAAGGTTTATTGGGAAAAGAGCC 1657
Qy 1739 CTTTAAATACGACAGCAACCTGGAGACACCGCTTTTTCATCTCATCTGCATGCCGAGCGGATG 1798
Db 1658 CATTAAGTACGCTCCCAACGTTGAACCTCCCTGCTTATATCCACTTACCGAAGACT 1717
Qy 1799 ACCGATGCCGATCGAGCAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATTTGGCAAGG 1858
```

1718 ACAGGTGTTGGCTCCCGAGGCAATTCGAACCTTCATATCCCTAAATATACCTGGGAAGA 1777
 1859 AAACCAAGCTGTCCGTTTCCGAAATCATCGCACAAATTTATCAGCAGCAGCAGCAGCAGC 1918
 1778 GAGTTGAATGGCAATATCCAGGAGAAATCATGACCTAAGTAGATCTCTGGGAAGCCAA 1837
 1919 GACAGGGGATCAAGCGCTGAAATATATATCATGCTCATGCTGTTGTGATCAACATCT 1970
 1838 AGCAGAGGTTAAAGACTTGAACATAATAGCAGATGGATGGAGAAATGGCT 1889

RESULT 5

ABQ86198

ID ABQ86198 standard; DNA; 1032 BP.

AC ABQ86198;

XX 09-SEP-2002 (first entry)

XX Lactobacillus rhamnosus HN001 polynucleotide SEQ ID NO 14.

DE Lactobacillus rhamnosus; strain HN001; vulnery; antilipemic; milk;
 KW immunostimulant; anti-infection; lactose digestion; immune system;
 KW metabolic activity; nutrition; health; transgenic; lactic acid bacteria;
 KW genome mapping; gastrointestinal disorder; dairy processing; vaccine;
 KW fermentation; probiotic; cholesterol; wound healing; gene; ds.

XX Lactobacillus rhamnosus.

XX WO200244383-A1.

XX 06-JUN-2002.

XX 28-NOV-2001; 2001WO-NZ00286.

XX 28-NOV-2000; 2000US-0724623.

XX (GENE-) GENESIS RES & DEV CORP LTD.

XX (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.

XX Glenn M, Havukkala IU, Lubbers MW, Dekker J;

XX WPI; 2002-519588/55.

XX P-PSDB; ABP61033.

XX Novel isolated Lactobacillus rhamnosus polynucleotides encoding
 PT polypeptide with anti-infection/lactose digestion modulating activity,
 PT useful to improve properties of microbes used in milk-derived products
 PT manufacture

XX Claim 2; Page 63; 128pp; English.

XX The invention relates to an isolated polynucleotide (I) comprising a
 CC nucleotide sequence (ABQ86185-ABQ86243) present in Lactobacillus
 CC rhamnosus strain HN001 that encodes a polypeptide (ABP61020-ABP61060)
 CC with activity such as enzyme activity; anti-infection activity; lactose
 CC digestion modulating activity; immune system modulating activity; amino
 CC acid, lipid, vitamin or carbohydrate metabolic activity, flavour, texture
 CC or aroma modulating activity. (I) is useful for improving the properties
 CC of microbes used in the manufacture of milk-derived products and
 CC probiotic supplements, which involves modulating the polynucleotide
 CC content or composition of the microbes by transforming the microbes with
 CC (I). (I) is also useful for identifying an organism (preferably a
 CC bacterial or yeast cell) or reproductive material or an extract from the
 CC organism, as having a specific origin. Proteins encoded by (I) are useful
 CC for modifying the flavour, aroma, texture and/or nutritional and health
 CC benefits of milk-derived products, which involves adding one or more
 CC polypeptides to the milk being processed. Genetic constructs comprising
 CC (I) are useful for modulating the polynucleotide content or composition
 CC of a organism. (I) is useful for identifying, isolating or synthesising
 CC DNA molecules such as promoter, DNA binding elements, open reading frames
 CC or full-length genes, that then can be used as expressible DNA in

CC transgenic organisms. (I) may be used to detect lactic acid bacteria,
 CC preferably *L. rhamnosus* in a sample material. (I) is also useful for
 CC genome mapping, physical mapping, and in positional cloning of genes of
 CC more or less related microbes, and to design probes and primers. (I) is
 CC also useful for transforming microbes for use in a therapeutic
 CC composition that is effective for treating or preventing a
 CC gastrointestinal condition or disorder caused by the presence of
 CC pathogenic microbes in the gastrointestinal tract or by the absence of
 CC normal intestinal microbes in the intestinal tract. Proteins are used to
 CC raise antibodies, to isolate corresponding interacting proteins, as
 CC nutritional additives and as additives in dairy processing and
 CC fermentation processing. (I) and encoded proteins are used for the
 CC selection and production of more effective probiotic bacteria, as
 CC bioactive (health promoting) ingredients and health supplements, for
 CC immune function enhancement; for reduction of blood lipids such as
 CC cholesterol; for production of bioactive material from genetically
 CC modified bacteria as adjuvants; for wound healing; in vaccine
 CC development, in selection and production of genetically modified rumen
 CC microorganisms for improved animal nutrition and productivity, better
 CC flavour and improved milk composition.

XX Sequence 1032 BP; 252 A; 220 C; 281 G; 279 T; 0 other;

Query Match.

9.9%; Score 195.6; DB 24; Length 1032;

Best Local Similarity 54.3%; Pred. No. 1.1e-51;

Matches 469; Conservative 0; Mismatches 379; Indels 16; Gaps 3;

QY 1050 CTTTCTCTTTCACCTGATGACAGCAGCATTATTATGCCAGTGTGACAAAGCCGGACAGCC 1109

DB 78 CTTTGCATTGACACCAGATCATCGTGGGTGGTGTTCCTACGTGAGATGACGATGACCATTC 137

QY 1110 GAGTGAGCTTTACAGTATCCCGCTTGGACAGGAGAGAAACAGCTACTGGCGGAATGA 1169

DB 138 GAGTCGGCTGGTTTATTTGATTTGGCATCGGAGAGACAGGTTTGTATGACCCGAA 197

QY 1170 CAAGTTTGTACGGGAGCATACGATATCAATACCTGAAGAGATTCAATGCTACAGAAGA 1229

DB 198 TCGTCAGGTAACACAGCTCACTTGGGCTTAGTTACCCCTCAACGCTTAATTTTCAACGAGA 257

QY 1230 CGGCGTGATGTTGAACGGCTGGCTGATGAGGCTGCACAAATGGAAGTGCAGCAACATA 1289

DB 258 CGGTTTGTAGATGAGGCTGGTATTTCCACCGCAACAGCGGTCA-----TCATCGCA 311

QY 1290 TCCACTTATTTTAAATACACAGCGGCTCCGATATGATGATGAGGATACATATATTTTCA 1349

DB 312 TCCGCAATTTTGTATGTCATCGGCGCCAGCAGTCGGATATGGCTATACCTTTTCCA 371

QY 1350 TGAGTTTCAAGTGTGGCGGGAAGGATACGCGGTCTTTATATCAATCCGAGAGGAAG 1409

DB 372 TGAATGCAATCTCTGGCAGCAAAAGGCTATGGCGTGATTTGTGCAATCCGCTGGAGG 431

QY 1410 CCACGGCTACGGGCAAGAAATTTGTGAATGCGGTGAGAGGATTTATGGGGAAGGATTA 1469

DB 432 GTTAGTTACCGGAGGCAATTTACGGGCGCTGCTAATTAACATTAACCGGCGCGGATTA 491

QY 1470 TGACGATGTGATGAGGCTGTGGATGAGGCTATCAAAACGAGATCCGATATGATCCTAA 1529

DB 492 TGAAGATTGCTTGGCTTCCGGTGAAGAAGCGCTAAAGCTCGATACAAACAATGATCCGCA 551

QY 1530 GGGGCTGGGTGTCACGGGCGGAGCTACGGAGGTTTTATGACCAACTGGATTCGCGGCA 1589

DB 552 AGCTATTTGTCACCTGGCGGTTCTTTATGGCGGGTTTATGACAACTGGATTTGACCCCA 611

QY 1590 GACGAACCGCTTTAAAGCTGCGCTTACCCAGCGCTCGATATCAAAATGGAATGATCCTTCA 1649

DB 612 TACGATCGTTTTAAAGCAGCGGTACCCAGCGTTCGATTTCCAAATGGCTGAGTATGTA 671

QY 1650 CGGCGCTGATGATGCGGCTATTCTTTACAGACTGGCAGCTTGA-----GCATGA 1700

DB 672 TGGTACCAGTGCATCGGTTATTACTTTTACCGTGGGAAGTGAAGAGTGGAGCTG 731

QY 1701 CATGTTTGAGGACACAGAAAGCTCTGGGACCGGCTCTCTTTAAATACCCAGCAACGT 1760

Db . 732 CGATTTGTCAGATGTGCAAGCGCTTTGGGATTTTTCACCATTTAGTCCATCATGATCATGC 791
QY 1761 GGAGACACCGCTTTTGTACTGATGCGGAGCGGGATGACCGATGCCGATC-GAGCAGG 1819
Db 792 CAGAACACCGACGCTTGTGTGATGACACAGTGAATGATGAACGCTGCCCATCGGCCAAG 851
QY 1820 CGGAGCAGCTGTTTATCGCTCTGAAAAAATGGCAAGAAACCAAGCTTGTCCGTTTTC 1879
Db 852 TAGAAAAGTTGATCATCGGTCTCAAACTGCATGCTGTTGAAACCAAGTTTCATCGGTTTC 911
QY 1880 CGAATGCATCGCACAAATTTATCAC 1903
Db 912 CAAAGTCAATCATGATTTGTCC 935

RESULT 6

ABK77575
ID ABK77575 standard; DNA; 477 BP.

XX ABK77575;

DT 13-AUG-2002 (first entry)

XX Bacillus clausii genomic sequence tag (GST) #418.

XX Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW physiological provocation; ds.

XX Bacillus clausii.

XX W0200229113-A2.

XX 11-APR-2002.

PF 05-OCT-2001; 2001WO-US31437.

XX 06-OCT-2000; 2000US-0680598.

PR 27-MAR-2001; 2001US-279526P.

XX (NOVO) NOVOZYMES BIOTECH INC.

PA (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;

XX WPI; 2002-416684/44.

XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -

XX Claim 11; SEQ ID NO 4866; 200pp; English.

XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus
CC genomic sequenced tags (GST), examining the array, and determining
CC relative gene expression by an observed hybridisation reporter signal of
CC a spot in the array. The method is useful for measuring the expression of
CC genes in a first Bacillus cell relative to expression of the same genes
CC in one or more second Bacillus cells. The method is useful for monitoring
CC global expression of several genes from a Bacillus cell, discovering new
CC genes, identifying possible functions of unknown open reading frames and
CC monitoring gene copy number variation and stability. Monitoring changes
CC in expression of genes may be used to provide a representation of the way
CC in which Bacillus cells adapt to changes in culture conditions,
CC environmental stress or other physiological provocation. Extensive
CC follow-up characterisation is unnecessary, when one spot on an array
CC equals one gene or one open reading frame, since sequence information is
CC available. This sequence represents a genomic sequence tag (GST) used in
CC the method of the invention.

CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at
CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 477 BP; 138 A; 89 C; 118 G; 132 T; 0 other;

Query Match 9.8%; Score 192.2; DB 24; Length 477;
Best Local Similarity 65.1%; Pred. No. 8.6e-51;
Matches 300; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

QY 1499 CTATCAAAACGAGATCCGCATATTGATCTTAAGCGCTCGGTGTACACGGCGGAGCTACG 1558
Db 7 CAATTGCAACAATAGCTGGATTGATCAGGAGCGCATCGGTCTGCTGAGGGAAGTTACG 66

QY 1559 GAGTTTTATGACCAACTGGATCTCGGGGAGAGCAACCGCTTTAAAGCTCCGTTACCC 1618
Db 67 GTGGCTTTATGACGAATTGGGCTCTTGCCCATAGCAATCGGTTCAAAGCGCGCTTACGC 126

QY 1619 AGCGCTCGATATCAAAATGGATCAGCTTTCACGGCGTCAGTGATATCGGCTATTTCTTTA 1678
Db 127 AAGCTCGATATCAAAATGGATTAGTTTATGGCGTAAGCGATATTGGCTACTATTTT 186

QY 1679 CAGACTGGCAGCTTGAGCATGACATGTTTGAGGACACAGAAAAGCTCTGGGACCGGCTC 1738

Db 187 CAGAGTGGCAAAATAAAGACAGAC---TTGCATGATATCGAAACACTGTGGCCCATTCG 243

QY 1739 CTTTAAATACGACGAAACGTGGAGACACCGCTTTTGATACTGCATGGCGAGCGGATG 1798

Db 244 CACTCAAAATACGTTGAAAAGCTGGAACGCGCATTTGCTTATTTCTGCACGTCGAAAAGATT 303

QY 1799 ACCGATCCCGATCGACGAGCGGACGAGCTGTTTATCGCTCTGAAAAAATGGCAAG 1858

Db 304 ATCGCTCCGATTTGAACAGCAGACGACCAATTTGTTATTCGTTGAAAAGCATGGGAAG 363

QY 1859 AAACCAAGCTTGTCGGTTTTCGGAATGCATCGCAATTTATCACGCCACCGACACCCAA 1918

Db 364 AGACTGTTATTTATCCGTTTCCCAAGCAACCATCATGAATCTCAAGAAGTGTAAACCTA 423

QY 1919 GACAGCGGATCAAGCGCTGAAATATATCATGCTCATGTTT 1959

Db 424 ATTTGGGATTGAACGCTCTCAATCGCATAGCAGATTGGTTT 464

RESULT 7

AAH41226/C
ID AAH41226 standard; DNA; 349980 BP.

XX AC AAH41226;

XX DT 29-OCT-2001 (first entry)

XX Pyrococcus abyssi genomic fragment #5.

DE Pyrococcus abyssi genomic fragment #5.

XX Hyperthermophilic archaeon; hyperthermophilic protein; ds.

XX Pyrococcus abyssi.

Key Location/Qualifiers
FT misc_feature 1..49980
FT /*tag= a
FT /note= "This sequence overlaps with the 3' end of
FT AAH41225"
FT misc_feature 300001..349980
FT /*tag= b
FT /note= "This sequence overlaps with the 5' end of
FT AAH41227"

PN FR2792651-A1.

XX 27-OCT-2000.

XX 21-APR-1999; 99FR-0005034.


```
XX 21-APR-1999; 99FR-0005034.
XX (CNRS ) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querrou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX
XX Claim 1; Page 511-606; 1657pp; French.
XX
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi and P. abyssi proteins (see AAB96053-AAB96842). P. abyssi is a
XX hyperthermophilic archaeon, which is isolated from deep-sea hydrothermal
XX vents. The present sequence is a fragment of the genomic sequence of P.
XX abyssi. The 5' end of this sequence overlaps with the 3' end of AAB41225
XX and the 3' end of this sequence overlaps with the 5' end of AAB41227. The
XX proteins of the present invention have various potential industrial uses,
XX since the proteins are stable at very high temperatures, some up to 110
XX degrees centigrade.
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAB75903-AAB75920 and AAB66436.
XX
XX Sequence 349980 BP; 98084 A; 80447 C; 77665 G; 93784 T; 0 other;
XX
XX Query Match 8.0%; Score 158; DB 22; Length 349980;
XX Best Local Similarity 52.7%; Pred. No. 5e-38;
XX Matches 391; Conservative 0; Mismatches 345; Indels 6; Gaps 2;
XX
QY 1230 CGCGGTGATGGTGAACGGCTGGCTGATGAGCGCTGCACAAATGAAGGTGAGACAATA 1289
DB 1230 CGCGGTGATGGTGAACGGCTGGCTGATGAGCGCTGCACAAATGAAGGTGAGACAATA 1289
QY 205109 CGACTTGGAGATAGACGGCTGGTACATATAAGCGGCTCAAGGAGGCGGAGAGGC--- 205053
DB 205109 CGACTTGGAGATAGACGGCTGGTACATATAAGCGGCTCAAGGAGGCGGAGAGGC--- 205053
QY 1290 TCCACTTATCTTAACATACACCGCGCTCGCATATGATGATGACGACATACATATTTCA 1349
DB 1290 TCCACTTATCTTAACATACACCGCGCTCGCATATGATGATGACGACATACATATTTCA 1349
QY 205052 TCCGGTAATAGTCTTCCTCCACCGGTGGGCGCCCAAGGGAATGTACGCTACTACTCAAGTA 204993
DB 205052 TCCGGTAATAGTCTTCCTCCACCGGTGGGCGCCCAAGGGAATGTACGCTACTACTCAAGTA 204993
QY 1350 TGAGTTTCAGGTGCTGGCGGCGGAAAGGATACGCGGTGCTTTATATCAATCCGAGGAGAG 1409
DB 1350 TGAGTTTCAGGTGCTGGCGGCGGAAAGGATACGCGGTGCTTTATATCAATCCGAGGAGAG 1409
QY 204992 CGAGATGCGACGTTATGCGGAGCAAGGGCTACTACATAGTCTATCTCAATCAAGGGAAG 204933
DB 204992 CGAGATGCGACGTTATGCGGAGCAAGGGCTACTACATAGTCTATCTCAATCAAGGGAAG 204933
QY 1410 CCACGGGTACGGCGGAGAAATTTGTAATGCGGTGATGAGGAGATTAATGGGGAAGGATTA 1469
DB 1410 CCACGGGTACGGCGGAGAAATTTGTAATGCGGTGATGAGGAGATTAATGGGGAAGGATTA 1469
QY 204932 CAACGGTTACAGCGGAGGACTTTGCACCTTAGAGTCTCTCGAGAGGACTGGCCTTAGAGGACTT 204873
DB 204932 CAACGGTTACAGCGGAGGACTTTGCACCTTAGAGTCTCTCGAGAGGACTGGCCTTAGAGGACTT 204873
QY 1470 TGACGATGATGACGCGCTGTGGATGAGGCTATCAAAACGAGATCCGCATATTGATCTTAA 1529
DB 1470 TGACGATGATGACGCGCTGTGGATGAGGCTATCAAAACGAGATCCGCATATTGATCTTAA 1529
QY 204872 CCAGGATATCTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 204813
DB 204872 CCAGGATATCTGAATGGAATAGAGGAGTTCTTCAAGCTTGAACCGCAAGCGGACAGGGA 204813
QY 1530 GCGGCTCGGTGTCACGGCGGCGGAAAGCTACGGAGGTTTTATGACCAACTGGATCGCGGCA 1589
DB 1530 GCGGCTCGGTGTCACGGCGGCGGAAAGCTACGGAGGTTTTATGACCAACTGGATCGCGGCA 1589
QY 204812 GAGGGTTGGAATAACGGGAATAAGCTACGGAGGCTTCATGACGAAGTGGGCGCTTAACGCA 204753
DB 204812 GAGGGTTGGAATAACGGGAATAAGCTACGGAGGCTTCATGACGAAGTGGGCGCTTAACGCA 204753
QY 1590 GACGAACCGCTTTAAGCTGCGGTACCCAGCGCTGATATCAATTTGATGATCAAGTTTCA 1649
DB 1590 GACGAACCGCTTTAAGCTGCGGTACCCAGCGCTGATATCAATTTGATGATCAAGTTTCA 1649
QY 204752 GAGCGACCTATTCAAGGCGCGGAATACGAGAGCGGAATAAGCTGACTGGCTTAACGAGTTA 204693
DB 204752 GAGCGACCTATTCAAGGCGCGGAATACGAGAGCGGAATAAGCTGACTGGCTTAACGAGTTA 204693
QY 1650 GCGGCTCAGTGATGCGCTATTCTTTTACAGACTGCGGCTTTCAGCATGATGATTTGA 1709
DB 1650 GCGGCTCAGTGATGCGCTATTCTTTTACAGACTGCGGCTTTCAGCATGATGATTTGA 1709
QY 204692 CCGGCTTCTCCGACATAGGATTTGGTTT---CGACAAGAGTTATAGTGAGAACCGCT 204636
DB 204692 CCGGCTTCTCCGACATAGGATTTGGTTT---CGACAAGAGTTATAGTGAGAACCGCT 204636
QY 1710 GGACACAGAAAGCTCTGGACCGGTCTCTTTAAAATACGACCAACGCTGGAGACAC 1769
DB 1710 GGACACAGAAAGCTCTGGACCGGTCTCTTTAAAATACGACCAACGCTGGAGACAC 1769
QY 204635 AGAGACGAGAACTTCAGGAAGCTAAGTCCGCTGTTCTACGCTAAGAGAGTGAAGGCTCC 204576
DB 204635 AGAGACGAGAACTTCAGGAAGCTAAGTCCGCTGTTCTACGCTAAGAGAGTGAAGGCTCC 204576
QY 1770 GCTTTTGAATGATGCGGCGGAGTATGACCGATGCCGATGCCGATCGAGAGCGGAGCAGCT 1829
DB 1770 GCTTTTGAATGATGCGGCGGAGTATGACCGATGCCGATGCCGATCGAGAGCGGAGCAGCT 1829
```

```
DB 204575 CCTGTGCTTAATCCACAGCCTTGAAGATTATCGCTGTCCATTGGATCAGAGCTTGTGATTT 204516
QY 1830 GTTTATCGCTCTGAAAAAATGGGCAAGGAAACCAAGCTGTGTCGGTTTTCCGAATGCATC 1889
DB 204515 CTACACAGCTCTGAAGATTAGGCAAGAAAGTTTACATAGCGATATTCAGAGGGGAGC 204456
QY 1890 GCACAATTTATCAGCAGCGGACACCCAAAGACAGCGATCAAGCGCTGAATTATATCAG 1949
DB 204455 TCATGACATAGCATAAAGGAGGCCCAAGACACAGGATGAAGAGTACAAGCTTTCAT 204396
QY 1950 CTCATGGTTTGTATCAACATCTC 1971
DB 204395 GGAGTTCTTCGAGAGGAAGCTC 204374
```

RESULT 8

```
AAV12887
ID AAV12887 standard; DNA; 1869 BP.
XX
AC AAV12887;
XX
DT 22-JUN-1998 (first entry)
XX
DE Thermococcus amidase gene.
XX
KW Amidase; thermostable enzyme; optically active compound;
KW L-amino acid; peptide; peptidomimetic; archaeobacterium; ss.
XX
OS Thermococcus sp. strain GU5L5.
XX
FH Key Location/Qualifiers
FT CDS 1..1869
FT /tag= a
FT /transl_except= (pos:1228..1230, aa:Val)
FT /transl_except= (pos:1231..1233, aa:Val)
XX
PN WO9748794-A1.
XX
PD 24-DEC-1997.
XX
XX 17-JUN-1997; 97WO-US09319.
XX
XX 17-JUN-1996; 96US-0664646.
XX
XX (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Murphy D, Reid JC, Robertson D;
XX
DR WPI; 1998-063135/06.
DR P-PSDB; AAW41248.
XX
XX Nucleic acid encoding heat-stable amidase from Thermococcus GU5L5 -
XX useful for removing N-terminal amino acids from synthetic peptide(s)
XX and peptidomimetic(s)
XX
XX Claim 2; Page 39-41; 56pp; English.
XX
XX This novel isolated nucleic acid encodes a 622 amino acid
XX thermostable amidase (see AAW41248) of the archaeobacterium
XX Thermococcus GU5L5. It was isolated from a genomic library of
XX positive clone to PCR amplification (see AAV12888-89). The nucleic
XX acid can be used in the production of amidase in host cells, and to
XX identify related sequences encoding similar enzymes. The encoded
XX amidase is used to remove Arg, Phe or Met from the N-terminus of
XX synthetic peptides or peptidomimetics (claimed). Removal of the
XX N-terminal residue can be done even in the presence of a more
XX reactive ester bond (very difficult to achieve non-enzymatically).
XX The amidase is selective for L-amino acids and can therefore be
XX used to produce optically active compounds. The protein tolerates
XX temperatures up to at least 70 degC and high concentrations of
XX organic solvent (e.g. over 40% dimethyl sulphoxide) so can cleave
XX bonds that are normally resistant.
```

```
XX
SQ Sequence 1869 BP; 476 A; 486 C; 567 G; 340 T; 0 other;

Query Match          7.6%; Score 149.8; DB 19; Length 1869;
Best Local Similarity 52.3%; Pred. No. 8.2e-37;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATCTTAACATACACGCGGTCGCGATATAGTACGGACATACATATTTTCAT 1350
Db 1135 CCGTGATAGTCTTCGTCACCGCGCGCGGATGATGATGATGATGATGATGATGATGAT 1194
QY 1351 GAGTTTCAGGTCGTCGCGCGGAAAGGATACCGGCTCGCTTTATATCAATCCGAGAGAAAGC 1410
Db 1195 GAGATGCAAGTATGCGGAGCAAGGCTACTACTGCTGCTCGTGAAACCGCGCGGACG 1254
QY 1411 CACGGCTACGGCAGGAATTTGTAATCGGTCGAGAGAGATATATGGGGAAAGATAT 1470
Db 1255 GACGGCTATACGAAGATTCGCGCTCGCGCTCGGAGAGAGACTGGCTTGGAGACTTT 1314
QY 1471 GACGATGTGATCGAGGCTGTGGATGAGGCTATCAACGAGATCGCATATTTGATCCTAAG 1530
Db 1315 GAGACATAATGAACGGCATCGAGAGTTCCTCAAGCTCGAACCGCGGACAGGGAG 1374
QY 1531 CGGCTCGGTGTACGCGCGGAAAGCTACGAGGTTTATGACCACTGGATCGTGGGCGAG 1590
Db 1375 CGCGTTGGAATAACGGGCATAAAGCTACGCGCGCTTCATGACCACTGGGCGCTTGACTCAG 1434
QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCCAGCGCTCGATATCAATTTGGATCAGCTTTCAC 1650
Db 1435 AGCGACCTCTTCAAGGAGGAATAAGCGGAGAACGGCATTAAGCTACTGCGCTCACCAGCTAC 1494
QY 1651 GCGCTCAGTATATCGGCTATTTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTCAG 1710
Db 1495 GCCTTCGACATAGGCTCTGTACGACGTGAGGTCTACGCGGCGCAATTCGGTTAGAG 1554
QY 1711 GACACGAAAGCTCTGGGACCGGTCTCCTTTTAAATACGCGCAAAAGCTGGAGACACCG 1770
Db 1555 AAC---GAGAACTTCAGGAAGCTCAGCGCGCTGTCTACGCTCAGAACGTGAAGCGCGCG 1611
QY 1771 CTTTGTGATGATGCGGAGCGGATGACCGGATGACCGGATGACCGGATGACCGGATGAC 1830
Db 1612 ATACTCTATCCACTCGCTTGAGGACTACCGCTGCGGCTCGACGAGCGCTTATGTTTC 1671
QY 1831 TTTATCGCTCTGAAATAAATGGCAAGGAACCAAGCTTGTCGCTTTTCCGAATGCATCG 1890
Db 1672 TACAACGTCTCAAGGACATGGGCAAGGACCTACATAGCATATTCACGCGCGCGCC 1731
QY 1891 CACAATTTATCAGCACCGGACACCCAGACAGCGGATCAAGCGCCTGAAATATATATCAGC 1950
Db 1732 CACGGCCACAGCGTCCGCGGAAGCCGAGGACAGCGGCGGAGCGCTACAGGCTCTTCATA 1791
QY 1951 TCATGGTTTGTATCAACATCTC 1971
Db 1792 GAGTCTTCGAGCGCAAGCTC 1812

RESULT 9
ID ABK77547
XX ABK77547 standard; DNA; 585 BP.
AC ABK77547;
XX
XX 13-AUG-2002 (first entry)
XX
XX Bacillus clausii genomic sequence tag (GST) #390.
XX
XX Differential gene expression; genomic sequenced tag; GST;
XX altered culture condition; environmental stress;
XX physiological provocation; ds.
OS Bacillus clausii.
XX
XX WO200229113-A2.
XX
```

```
XX
PD 11-APR-2002.
XX
XX 05-OCT-2001; 2001WO-US31437.
XX
XX 06-OCT-2000; 2000US-0680598.
XX
XX 27-MAR-2001; 2001US-279526P.
XX
XX (NOVO ) NOVOZYMES BIOTECH INC.
XX (NOVO ) NOVOZYMES AS.
XX
XX Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.
XX
XX Monitoring differential expression of several genes in first Bacillus
XX cell relative to expression of same genes in one or more second
XX Bacillus cells, by using substrate containing Bacillus genomic
XX sequenced tag array -
XX
XX Claim 11; SEQ ID NO 4838; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
XX genes in a first Bacillus cell relative to expression of the genes in
XX other Bacillus cells, comprising hybridising labelled nucleic acid probes
XX isolated from Bacillus cells to a substrate containing array of Bacillus
XX genomic sequenced tags (GST), examining the array, and determining
XX relative gene expression by an observed hybridisation reporter signal of
XX a spot in the array. The method is useful for measuring the expression of
XX genes in a first Bacillus cell relative to expression of the same genes
XX in one or more second Bacillus cells. The method is useful for monitoring
XX global expression of several genes from a Bacillus cell, discovering new
XX genes, identifying possible functions of unknown open reading frames and
XX monitoring gene copy number variation and stability. Monitoring changes
XX in expression of genes may be used to provide a representation of the way
XX in which Bacillus cells adapt to changes in culture conditions,
XX environmental stress or other physiological provocation. Extensive
XX follow-up characterisation is unnecessary, when one spot on an array
XX equals one gene or one open reading frame, since sequence information is
XX available. This sequence represents a genomic sequence tag (GST) used in
XX the method of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at
XX ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 585 BP; 173 A; 109 C; 142 G; 160 T; 1 other;

Query Match          4.2%; Score 82.4; DB 24; Length 585;
Best Local Similarity 48.0%; Pred. No. 1.7e-15;
Matches 233; Conservative 0; Mismatches 252; Indels 0; Gaps 0;

QY 587 TTTCTCCTGACGGCAAAATGGCTGTGTTTCTCAGCTAAATTTAACTGAAACAGATGATGCA 646
Db 1 TTTGAGCAAAATGGACGTCTGTTCTGTAGCGGCGAGCCATAAGGAAAAACAATGATTTTA 60
QY 647 GCAAGCCCATGATGTTTACATAATGTCACCTGAGTCTGGAGATCTTAAGCAGGTTACAC 706
Db 61 CTTTTCAGTGTGTTTGGTTTATTTGATTCGGGAGCAAAAACGCAAAATACAA 120
QY 707 CTCATCGCGGCTCATTCGGATCAAGCTCATTTTCCACAGACGCAAGGATCTGCTTTGCG 766
Db 121 AGGAACCAATGTCGTACAAAGCGGATCTGTATCACCAGATGTTCTTATGCTGCTTCC 180
QY 767 TTGGAATGAAAGGAATATAGAAATGCTAGCTCTCAAGAGCGTGGCTCTATGATATCG 826
Db 181 TAGGCCATGAGAGAGAAATACAAAACGCGACATTGACGAAGCTGTATGTTTATGATGCC 240
QY 827 AACAGGCGGCTCAGATGTTCTTACTGAGATGCTGAGCTTCATTACCGGATCGCTGA 886
Db 241 AAAACAAGAAATCCGTTTGTCTTCCGGCCCAAGTGGATGCGTCAATTCGGCATGTTGCAA 300
QY 887 TTGGAGATTCATTGATCGTGTGTAACAGCGCCCGCTTTTGGACAAAGGACAGCCCAAG 946
```

Db 301 TTGGGATTTTACAGATATTTGGTTCAAAATGGTTGAAATGGGAGAGATGGTCAGT 360
QY 947 GGTATTTATGTCATCGGCACAGATCAAGGAGTACGGGCATCTATTATATTTTCGATTGAAG 1006
Db 361 CCATTTATGCGCTTGTTCAAAAGAGGCGATGTAATGTTGGTCGTTTATAGCCTCGATG 420
QY 1007 GCCTTGTATCGGATTCGTCGTAAGAAAGAGTACATCAATAGCTTTTCTTTTCACCTG 1066
Db 421 GCACATCAATCAGTCACGCATGAAAGAGTCAATAATGGATTCGATTTTCAAAATG 480
QY 1067 ATGAA 1071
Db 481 GCCAA 485

RESULT 10

AAFI13721
ID AAF13721 standard; cdna; 659 BP.

AC AAF13721;
DT 13-MAR-2001 (first entry)

DE Aspergillus oryzae EST SEQ ID NO:6244.
XX Multiple gene expression; filamentous fungal cell; EST;
XX expressed sequence tag; Fusarium venenatum; Aspergillus niger;
KW Aspergillus oryzae; Trichoderma reesei; identification; recombination;
KW culture condition; environmental stress; spore morphogenesis;
KW metabolic pathway engineering; catabolic pathway engineering; ss.

OS Aspergillus oryzae.
XX
XX WO200056762-A2.
XX
XX 28-SEP-2000.
XX
XX 22-MAR-2000; 2000WO-US07781.
XX
XX 22-MAR-1999; 99US-0273623.
XX
XX (NOVO) NOVO NORDISK BIOTECH INC.
PA (NOVO) NOVO NORDISK AS.

XX Berka RM, Rey MW, Shuster JR, Kauppinen S, Clausen IG, Olsen PB;
XX WPI; 2000-594572/56.
XX
XX Monitoring differential expression of genes in filamentous fungal cells
PT uses fluorescence-labeled nucleic acids isolated from the cells and a
PT substrate of expressed sequence tags -
XX
XX Claim 88; Page 2569; 3161pp; English.

XX The present invention describes a method for monitoring differential
CC expression of genes in a first filamentous fungal (FF) cell relative to
CC expression of the same genes in one or more second filamentous fungal
CC cells. The method uses fluorescence-labeled nucleic acids isolated from
CC the FF cells and a substrate of expressed sequence tags (EST). The ESTs
CC are used in the methods for monitoring differential expression of genes
CC in a first filamentous fungal (FF) cell relative to expression of the
CC same genes in one or more second filamentous fungal cells. Monitoring
CC the global expression of genes from FF cells allows the production
CC potential of the microorganisms to be improved. New genes may be
CC discovered, possible functions of unknown open reading frames can be
CC identified and gene copy number variation and stability can be
CC monitored. The expression of genes can be used to study how FF cells
CC adapt to changes in culture conditions, environmental stress, spore
CC morphogenesis, recombination, metabolic or catabolic pathway
CC engineering. Using ESTs provides several advantages over genomic or
CC random cDNA clones including elimination of redundancy as one spot on an
CC array equals one gene or open reading frame, and organisation of the

CC microarrays based on function of the gene products to facilitate
CC analysis of the results. AAF07478 to AAF11247 represents ESTs from
CC Fusarium venenatum; AAF11248 to AAF11853 represents ESTs from Aspergillus
CC niger; AAF11854 to AAF14878 represents ESTs from Aspergillus oryzae; and
CC AAF14879 to AAF15337 represents ESTs from Trichoderma reesei, which are
CC all specifically claimed in the present invention.

XX Sequence 659 BP; 162 A; 214 C; 157 G; 126 T; 0 other;

Query Match 2.5%; Score 49.8; DB 21; Length 659;
Best Local Similarity 51.1%; Pred. No. 5.4e-05;
Matches 117; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 1358 AGGTGCTGGCGCGAAGGATACGCGGTGCTTTATATATCAATCCGAGAGGAGCCACGGCT 1417
Db 47 AGGTCTTCGCCGACCCAGGATACGTTAGTCGTCGACCCCAACCCGGAAGCAGTGGCT 106
QY 1418 ACGGCGAGGAATTTGTAATCGGTCAGAGGAGATTTATGGGGGAAGGATTTATGACCATG 1477
Db 107 TCGGCGATGCCCTCCAGGACGCCATCCAAACCAATGGGAGGCTACCCCTACGAGACC 166
QY 1478 TGATGCAGGCTGTGGATGAGGTATCAAAACGAGATCGGCATATTGATCTTAAGCGGCTCG 1537
Db 167 TCGTCAAGGCTGGGAATACGTCACGAGAACTTCGACTTCATTGACACGACACCGGTG 226
QY 1538 GTGTACGCGGCGAAGCTACGAGGTTTTATGACCAACTGGATCGTGG 1586
Db 227 TCGCGCGCGGTGCCAGCTACGGTGGCTTCATGATCAACTGGATCCAAGG 275

RESULT 11

ABK73047
ID ABK73047 standard; DNA; 1797 BP.

XX
XX AC ABK73047;
XX
XX DT 13-AUG-2002 (first entry)

XX DE Bacillus licheniformis genomic sequence tag (GST) #338.
XX KW Differential gene expression; genomic sequenced tag; GST;
KW altered culture condition; environmental stress;
KW Physiological provocation; ds.

XX OS Bacillus licheniformis.

XX
XX PN WO200229113-A2.
XX
XX PD 11-APR-2002.

XX
XX PF 05-OCT-2001; 2001WO-US31437.
XX
XX PR 06-OCT-2000; 2000US-0680598.
XX
XX PR 27-MAR-2001; 2001US-279526P.

XX
XX PA (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.

XX Berka R, Clausen IG;
XX
XX WPI; 2002-416684/44.

XX
XX Monitoring differential expression of several genes in first Bacillus
PT cell relative to expression of same genes in one or more second
PT Bacillus cells, by using substrate containing Bacillus genomic
PT sequenced tag array -

XX
XX Claim 4; SEQ ID NO 338; 200pp; English.
XX
XX The invention describes a method of monitoring differential expression of
CC genes in a first Bacillus cell relative to expression of the genes in
CC other Bacillus cells, comprising hybridising labelled nucleic acid probes
CC isolated from Bacillus cells to a substrate containing array of Bacillus

methylation of a particular cytosine in a motif 5'-CpG-3', present in a genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridised to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridisation to both classes is determined from the label on the amplicon. From the ratio of labels hybridised to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. The ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 788 BP; 91 A; 86 C; 291 G; 320 T: 0 other;

Query Match 1.98; Score 36.6; DB 24; Length 788;

Best Local Similarity 49.2%; Pred. No. 1.1;

Matches	96;	Conservative	0;	Mismatches	99;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	-----	--------	----	------	----

Qy	49	ACCGATCCTCAATACGCGCCACAGCGGTACCGGTGCCGCATATGTAAATATCAACAATAAT	108
Db	264	AACGAAACGGCGCGACGCCCAATAACAACATCGACCGTCATTACTCAAAATCAACAATAAA	205
Qy	109	CAAGAGAAAGATTTCGTATACATCAAAATATATGGATCTATGAAACGAAACGGAGGATCT	168
Db	204	AAACTCGAAACCGGTACGAAACAATAATACCGAAACTACGAAAAAAAACCGAAACCGAA	145
Qy	169	GTTCCTTTGGACACATGGAGAAAAAGCGAACCGACCGACCATGGTCTTCGGACCGGGCGC	228
Db	144	ACGCCGTAAAAAATAAAAAAATAAAAAACCGAAATCTTCCCAAAAAAACTCGCGCGCGCCGA	85
Qy	229	ACGCTTGCTTTATT	243
Db	84	ACGACCGCTTAACT	70

Search completed: July 4, 2003, 13:20:01
Job time : 458 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 02:26:30 ; Search time 120 seconds
(without alignments)
5037.166 Million cell updates/sec

Title: US-09-462-845-1
Perfect score: 1971
Sequence: 1 atgaaaagtgataaccgc.....catggtttgatcaacatctc 1971

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCFUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	227.2	11.5	1896	3	US-09-016-080-2
2	149.8	7.6	1869	2	US-08-664-646A-1
3	149.8	7.6	1869	2	US-09-066-285-1
4	149.8	7.6	1869	3	US-09-261-006-1
5	149.8	7.6	1869	3	US-08-951-088-1
6	149.8	7.6	1869	4	US-09-609-566-1
7	149.8	7.6	1869	4	US-09-609-570-1
8	54.4	2.8	7218	1	US-08-232-463-14
9	39.4	2.0	3085	4	US-09-221-017B-1010
10	35.8	1.8	1395	3	US-08-834-776A-1
11	35.6	1.8	289	4	US-09-007-005-17
12	35.6	1.8	289	4	US-09-244-796-17
13	34.2	1.7	729	2	US-08-512-955-3
14	33.6	1.7	3274	4	US-09-221-017B-661
15	33.6	1.7	6152	1	US-08-557-139-1
16	33.4	1.7	957	4	US-09-601-027-2
17	33.4	1.7	2277	1	US-08-676-967-2
18	33.4	1.7	2277	1	US-08-676-974-2
19	33.4	1.7	2277	1	US-09-098-487-2
20	32.4	1.6	7680	4	US-09-210-748A-3
21	31.8	1.6	957	1	US-07-688-299-2
22	31.8	1.6	957	1	US-07-980-517A-2
23	31.8	1.6	2046	1	US-07-688-299-12
24	31.8	1.6	2046	1	US-07-980-517A-11
25	31.4	1.6	2743	1	US-08-396-479B-3
26	31.4	1.6	2743	1	US-08-818-823-3
27	31.4	1.6	2749	2	US-08-124-981A-1

```

c 28 31.4 1.6 2751 3 US-09-037-190-45 Sequence 45, Appl
c 29 31.4 1.6 2751 3 US-09-037-192-45 Sequence 45, Appl
c 30 31.4 1.6 2751 3 US-09-037-143-45 Sequence 45, Appl
c 31 31.4 1.6 2751 4 US-09-049-691-45 Sequence 45, Appl
c 32 31.4 1.6 2751 4 US-08-260-174-45 Sequence 45, Appl
c 33 31.4 1.6 2751 4 US-09-338-128A-45 Sequence 45, Appl
c 34 31.4 1.6 2751 4 US-09-037-192-45 Sequence 36, Appl
c 35 31.4 1.6 2853 5 PCT-US94-07297-36 Sequence 45, Appl
c 36 31.4 1.6 5502 4 US-09-232-346-45 Sequence 45, Appl
c 37 31.2 1.6 394 4 US-09-221-017B-49 Sequence 45, Appl
c 38 31.2 1.6 504 4 US-09-134-001C-1816 Sequence 1816, Ap
c 39 31.2 1.6 1308 4 US-09-134-001C-1785 Sequence 1785, Ap
c 40 31.2 1.6 1348 2 US-08-743-637B-15 Sequence 15, Appl
c 41 31.2 1.6 1348 3 US-08-526-840B-15 Sequence 15, Appl
c 42 31.2 1.6 1607 4 US-08-540-650B-17 Sequence 17, Appl
c 43 31.2 1.6 4928 1 US-08-399-561-1 Sequence 1, Appl
c 44 31 1.6 1193 1 US-07-956-697B-4 Sequence 4, Appl
c 45 31 1.6 1193 1 US-08-263-098-4 Sequence 4, Appl

```

ALIGNMENTS

RESULT 1

US-09-016-080-2
; Sequence 2, Application US/09016080
; Patent No. 6133012

GENERAL INFORMATION:

; APPLICANT: Ishikawa, Kazuhiko
; APPLICANT: Matsui, Ikuo
; APPLICANT: Ishida, Hiroyasu
; APPLICANT: Kosugi, Yoshisugu
; APPLICANT: Higuchi, Kazuhiko
; TITLE OF INVENTION: THERMOSTABLE ACYL PEPTIDE HYDROLASE AND GENE ENCODING
; TITLE OF INVENTION: THE SAME
; FILE REFERENCE: 07898/022001
; CURRENT APPLICATION NUMBER: US/09/016,080
; CURRENT FILING DATE: 1998-01-30
; EARLIER APPLICATION NUMBER: JAPAN 18381/1997
; EARLIER FILING DATE: 1997-01-31
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 2
; LENGTH: 1896
; TYPE: DNA
; ORGANISM: Pyrococcus horikoshii
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1896)
US-09-016-080-2

Query Match 11.5%; Score 227.2; DB 3; Length 1896;
Best Local Similarity 55.4%; Pred. No. 1.5e-66;
Matches 461; Conservative 0; Mismatches 368; Indels 3; Gaps 1;

```

Qy 1139 AGGAAGAGAAACAGCTGCGCGGAATGACAACTTTGTTCAGGAGCATACGATATCAA 1198
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1061 ATGGAAGAGAGAGAGAGGTTACCGACTTTTACAAATGATAAGGTTTACACCCCTTCAA 1120
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1199 TACCTGAAGAGATTCAATATGCTACAGAAAGCGCGGTGATGGTGAACGGCTGGGTGATGA 1258
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1121 AACCTGAACACTTTTAAAGGTTTAAAGCAAGTGACGGGGTTGAAATAGATGCTGGGTAATGA 1180
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1259 GGCCTGCACAAATGGAAGGTGAGACACATATCCACTTATTCTTAAACATACACGCGGCTC 1318
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1181 AACCGGTGAACCTCAGGAAGAAAGATCCAGCTATTCTTAGAGATCCACGGTGGCTC 1240
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1319 CGCATATGATGATGCGGACATACATATTTTCATGAGTTTCAGGTGCTGCGCGGAAGGAT 1378
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1241 CTAAACCGCTTACGGTTACGCTTTTATGACAGAGTCCACGTTTAACTCTTAAGGCT 1300
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1379 ACCGGTGGTTTATATCAATCCGAGAGAACCCACGGCTACGGCAGGAATTTGTGAATG 1438
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 1301 TCCTCGTGATATTCTCAATCTTAGAGGCGGATGGCTACGGAGGAGTTCCGGGAT- 1359
QY 1439 CGGTGAGGAGATATTGGGGAAAGGATATGACGATGTGATGACGCTGTGATGAGG 1498
Db 1360 --ATAAGGGACATATGGGGAGAGGATACCAAGGATTTAATGAGGTAGTCGATGAAG 1417
QY 1499 CTATCAACAGATCCGCATATGATCCTAAGCGCTCGGTGTCACGGCGGAAGCTAGG 1558
Db 1418 CATTAAGGAGATTGACTTCATAGATGGGAAAGCTAGGAGTTACCGGGGGTTCCTATG 1477
QY 1559 GAGGTTTTATGACCAACTGGATCGTCGGGACAGAAACCGCTTTAAAGCTGCCGTTACCC 1618
Db 1478 GTGCTTCATGACGAAGTGGATAGTCGACATACCAACAGGTTCAAAAGCGCTGAACCC 1537
QY 1619 AGCGCTCGATATCAAAATGATCAGCTTTACGGGTCAGTGATATCGGCTATTCCTTTA 1678
Db 1538 AGAGATCAATTTCAAAATGGATAAGCTTCTCGGGACAAACGATATAGGTTATTACTTTG 1597
QY 1679 CAGACTGGCAGCTTGAGCATGACATGTTTGGAGGACACAGAAAGCTCTGGGACCGGTCTC 1738
Db 1598 CTCAGATCAATAGGAAAGATCCCTGGAGCACTTGGAGGTTATTGGGAAAGAGCC 1657
QY 1739 CTTTAAATAGCAGCAACGCTGGAGACACCGCTTTTATGATGATGCGGAGCGGATG 1798
Db 1658 CATTAAGTACGCTCCCAACGTTGAAACTCCCTGCTTATATCCATCCACTACCGAAGACT 1717
QY 1799 ACCGATCCCGATCGAGCGCGGAGCGAGCTGTTATCGCTCGTGAAGAAATGGCAAGG 1858
Db 1718 ACAGGTGTGCTTCCCGAGGCAATGCAACTCTTCATATCCCTAAATACCTGGGGAAGA 1777
QY 1859 AAACCAAGCTTGTCGGTTTTCCGAATGCATCGCACAAATTTATACGCGACCGGACACCCAA 1918
Db 1778 GAGTTGAATTGGCAATATCCAGGAGAAATCATGACCTAAGTAGATCTGGGAAGCCAA 1837
QY 1919 GACAGCGGATCAAGCGCTGAATTTATATCAGCTCATGTTTGTATCAACATCT 1970
Db 1838 AGCAGAGGTTAAAGACTTGAACCTAATAGCAGGATGATGGAGAAATGGCT 1889

RESULT 2

US-08-664-646A-1

; Sequence 1, Application US/08664646A

; Patent No. 5877001

; GENERAL INFORMATION:

; APPLICANT: Murphy et al.

; TITLE OF INVENTION: Amidases

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,

; STREET: 6 BECKER FARM ROAD

; CITY: ROSELAND

; STATE: NEW JERSEY

; COUNTRY: USA

; ZIP: 07068

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 3.5 INCH DISKETTE

; COMPUTER: IBM PS/2

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: WORD PERFECT 5.1

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/664.646A

; FILING DATE: June 17, 1996

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Charles J. Herron

; REGISTRATION NUMBER: 28.019

; REFERENCE/DOCKET NUMBER: 311400-53

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201-994-1700

; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
US-08-664-646A-1

Query Match 7.6%; Score 149.8; DB 2; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACITATTCTTAACATACACGGGCTCCGCATATGATGACGACATACATATTTTCAT 1350
Db 1135 CCGGTGATAGTCTTCCTCCACGGCGGCGCAAGGGCATGTACGACACCCGCTTGTCTAC 1194
QY 1351 GAGTTTCAGGTGCTGGCGGCAAGGATACCGGTCGCTTTATATCAATCCGAGAGGAAGC 1410
Db 1195 GAGATCAGCTGATGCGGAGCAAGGGCTACTACTGCTGCTCGTGAACCCGCGGCGAGC 1254
QY 1411 CACGGCTACGGGACAGAAATTTGTAATGCGGTGACAGGAGATTTATGGGAAAGGATTAT 1470
Db 1255 GACGGCTATAGCGAAGACTTTCGCGCTCCGCTGCTGGAGAGACTGGCTTGGAGACTTT 1314
QY 1471 GACGATGTGATGAGGCTGTGGATGAGGCTATCAACGAGATCCGCATATTTGATCCTAAG 1530
Db 1315 GAGGACATATGACGCGCATCGAGGAGTTCTTCAAGCTCGAACCAGGCGGACAGGAG 1374
QY 1531 CGGCTCGGTGTACGGGCGGAAGTACGAGGTTTTATGACCAACTGGATCGTGGGCGAG 1590
Db 1375 CGGCTTGAATAACGGGCATAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTGACTCAG 1434
QY 1591 ACCAACCGCTTTAAAGCTGCCGTACCCAGCGCTCGATATCAAAATTTGATCAGCTTTTCCAC 1650
Db 1435 AGCGACTCTTCAAGGCGAGGAATAAGCGAAGCAAGGATAGCTACTGCTCACCAGCTAC 1494
QY 1651 GGCGTCAGTGATATCGGCTATTTCTTTACAGACTGGCAGCTTGAGCATGACATGTTTGAG 1710
Db 1495 GCCTTCGACATAGGGCTCTGTACGAGCTCGAGGTTCATCGGGCCAAATCCGTTAGAG 1554
QY 1711 GACACAGAAAGCTCTGGGACCGGTCTTCTTTAAATACGCAAGCAAGCTGGAGACACCG 1770
Db 1555 AAC--GAGAACTTCAGAAAGCTCAGCCCGCTGTTCTACGCTCAGAACGCTGAAGCGCGC 1611
QY 1771 CTTTGTGATCTGATGCGGAGCGGATGACCGATGCCGATGCCGATCGAGCGGAGCAGCTG 1830
Db 1612 ATACTCCTAATCCACTCGCTTGAGGACTACCGCTGTCCGCTCGACAGAGCTTATGTTTC 1671
QY 1831 TTTATCGCTCTGAAAAAATGGGCAAGAAACCAAGCTTGTCCGTTTTCGAAATGCATCG 1890
Db 1672 TACAAGCTGCTCAAGGACATGGGCAAGAGCCCTACATAGGATATTCAAGCGCGCGCC 1731
QY 1891 CACAATTTATCAGCGACCGGACACCCCAAGACAGCGGATCAAGCGGCTCAATATATCAGC 1950
Db 1732 CACGGCCACAGCTCCCGGAAGCCGAGGACACAGCGGCGGAGCGCTTCTTCATA 1791
QY 1951 TCATGTTTTCATCAACATCTC 1971
Db 1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 3

US-09-066-285-1

; Sequence 1, Application US/09066285

; Patent No. 5985646

; GENERAL INFORMATION:

; APPLICANT: Murphy et al.

; TITLE OF INVENTION: Amidases

; NUMBER OF SEQUENCES: 4

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,


```
1471 GACGATGTGATGAGGCTGTGGATGAGGCTATCAACGAGATCGCATATATTGATCCCTAAG 1530
1315 GAGGACATATGAACGCATCGAGGATTTCTTCAAGCTCGAACCGAGCGGCGAGGAG 1374
1531 CGGCTCGGTGTCACGGCGGGAAGCTACGAGGTTTATGACCAACTGATCGTCGGGCGAG 1590
1375 CGGTTGGAATAACGGGCAATAAGCTACGCGGCTTCATGACCAACTGGGCGCTTGACTCAG 1434
1591 ACGAACGGCTTTAAAGCTGCGGTTTACCAGGCTCGATATCAAAATGATGATGATTCAC 1650
1435 AGCGACCTCTTCAAGGAGGAATAAGCGAGAACGCGCATAGGCTACTGGCTCACCAGCTAC 1494
1651 GCGGTGATGATATCGGCTATTTCTTACAGACTGGCAGCTTGAGCATGACATGATTTGAG 1710
1495 GCCTTCTCGGACATAGGCTCTGTTAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCG 1554
1711 GACACGAAAGCTCTGGGACCGCTCTCTTTTAAATACGACGACAAACGTCGAGACACCG 1770
1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTTCTACGCTCAGAGGCTTATGTTTC 1611
1771 CTTTGTGATGATGAGGCTGTGGGACCGGCTTGTCCGTTTTCGAAATGATCG 1830
1612 ATACTCTTAATCCACTCGCTTGAGGCTACCGGCTTCTACGCTCAGAGGCTTATGTTTC 1671
1831 TTTTGTGATGATGAGGCTGTGGGACCGGCTTGTCCGTTTTCGAAATGATCG 1890
1732 CAGGCGCACAGCTCGCGGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCG 1791
1951 TCATGTTTGTATCAACATCTC 1971
1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 5
US-09-51-088-1
; Sequence 1, Application US/08951088
; Patent No. 6136583
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; COMPUTER: IBM PS/2
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/664,646
; FILING DATE: June 17, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
US-08-951-088-1

Query Match 7.6%; Score 149.8; DB 3; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTTATCTTAAACATACACGCGGTCGCGATATGATGTGACGACATATATTTTCAT 1350
Db 1135 CCGGTGATAGTTCTTCTCCACGCGGCGGAGGCGATGTACGACACCGCTTCTGCTTAC 1194
QY 1351 GAGTTTCAGTGTGCGGCGGAAAGATATACGCGGTCGTTTATATCAATCCGAGAGGAGC 1410
Db 1195 GAGATGACGCTGATGCGGAGCAAGGCTACTACTGCTGCTTCGTGAACCCGCGCGGAGC 1254
QY 1411 CACGGCTACGGGCGAGGAATTTGTGAATGCGGTCGAGGAGATTTATGGGGAAGATAT 1470
Db 1255 GACGGCTATAGCAAGACTTCGCGCTCCGCGTCTCGAGAGGAGTGGCTTGGAGGACTT 1314
QY 1471 GACGATGTGATGAGGCTGTGGATGAGGCTATCAACGAGATCGGCATATTTGATCCTAAG 1530
Db 1315 GAGGACATATGAACGCGCATCGAGGAGTTCTTCAAGCTCGAACCGCGAGCGGAGGAG 1374
QY 1531 CGGCTCGGTGTCACGGCGGGAAGCTACGAGGTTTATGACCAACTGATCGTCGGGCGAG 1590
Db 1375 CGGTTGGAATAACGGGCAATAAGCTACGCGGCTTCTACGCAACTGAGGCTTATGTTTC 1434
QY 1591 ACGAACGGCTTTAAAGCTGCGGTTTACCAGGCTCGATATCAAAATGATGATGATTCAC 1650
Db 1435 AGCGACCTCTTCAAGGAGGAATAAGCGAGAACGCGCATAGGCTACTGGCTCACCAGCTAC 1494
QY 1651 GCGGTGATGATATCGGCTATTTCTTACAGACTGGCAGCTTGAGCATGACATGATTTGAG 1710
Db 1495 GCCTTCTCGGACATAGGCTCTGTTAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCG 1554
QY 1711 GACACGAAAGCTCTGGGACCGCTCTCTTTTAAATACGACGACAAACGTCGAGACACCG 1770
Db 1555 AAC---GAGAACTTCAGGAAGCTCAGCCCGCTTCTACGCTCAGAGGCTTATGTTTC 1611
QY 1771 CTTTGTGATGATGAGGCTGTGGGACCGGCTTGTCCGTTTTCGAAATGATCG 1830
Db 1612 ATACTCTTAATCCACTCGCTTGAGGCTACCGGCTTCTACGCTCAGAGGCTTATGTTTC 1671
QY 1831 TTTTGTGATGATGAGGCTGTGGGACCGGCTTGTCCGTTTTCGAAATGATCG 1890
Db 1672 TACAGCTGCTCAAGGACATGGGCAAGGAGGCTTACATAGCGATATTCAGGCGGCGCC 1731
QY 1891 CACAATTTATCAGCGACCGGACACCCAGAGGAGGATCAAGCGGCTGAATTTATATCAGC 1950
Db 1732 CAGGCGCACAGCTCGCGGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCGAGGCTCG 1791
QY 1951 TCATGTTTGTATCAACATCTC 1971
Db 1792 GAGTTCTTCGAGCGCAAGCTC 1812

RESULT 6
US-09-609-566-1
; Sequence 1, Application US/09609566
; Patent No. 6429004
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amides
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; ADDRESSEE: CECCHI, STEWART & OLSTEIN
```

```

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,566
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; US-09-609-566-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGACGACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGCGGAGGATGACGACACCGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGTGTGGCGGGAAGGATACGGGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGGCGGAGCAAGGCTACTACTGCTCTGCTGTAACCCGCGGCGCAG 1254

QY 1411 CACGGCTACGGGAGGATTTGTAATCGGTTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGGCTATACGGAAGACTTCGCGCTCGCGTCTCGGAGAGGACTGGCTTGGAGGACTTT 1314

QY 1471 GACGATGTGATGACGGTGTGGATGAGGTATCAACGAGATCGGCATATTTGATCCTAAG 1530
DB 1315 GAGGACATATGACGGCATCGAGGAGTCTTCAAGCTCGAACCGCGGCGGAGGAG 1374

QY 1531 CGGTCGGTGTACGGGCGGAGTACGGAGGTTTATGACCAACTGGATCGTGGGCGAG 1590
DB 1375 CGCGTTGGAATAACGGGCATTAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTGA 1434

QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCGAGGCTCGATATCAAAATTTGATCAGCTTTTAC 1650
DB 1435 AGCGACCTTTCAAGCAGGAGGATTAAGCAGAGAACGGCATTAAGCTTCTGCTCAGCTAC 1494

QY 1651 GCGCTCAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGGAGCATGACATGTTTGA 1710
DB 1495 GCCTTCGCGACATAGGCTCTGTTGACGACGCTCGAGGTCATCGGGCCAAATCCGTTAG 1554

QY 1711 GACACAGAAAGCTCTGGGACCGGTCTCTTTTAAATACGAGCAAAAGCTGGAGACACCG 1770
DB 1555 AAC---GAGAATTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAAGCTGAAGGCGCGC 1611

QY 1771 CTTTGTACTGATGGGACGGGATGACCGGATCGCGATCGGAGGCGGAGGAGGAGCTG 1830
DB 1612 ATACTCCTTAATCCACTCGCTTGGAGGACTTACCGCTGTCCGCTCGCAGGAGGAGCTTT 1671

; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,566
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; US-09-609-566-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGACGACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGCGGAGGATGACGACACCGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGTGTGGCGGGAAGGATACGGGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGGCGGAGCAAGGCTACTACTGCTCTGCTGTAACCCGCGGCGCAG 1254

QY 1411 CACGGCTACGGGAGGATTTGTAATCGGTTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGGCTATACGGAAGACTTCGCGCTCGCGTCTCGGAGAGGACTGGCTTGGAGGACTTT 1314

QY 1471 GACGATGTGATGACGGTGTGGATGAGGTATCAACGAGATCGGCATATTTGATCCTAAG 1530
DB 1315 GAGGACATATGACGGCATCGAGGAGTCTTCAAGCTCGAACCGCGGCGGAGGAG 1374

QY 1531 CGGTCGGTGTACGGGCGGAGTACGGAGGTTTATGACCAACTGGATCGTGGGCGAG 1590
DB 1375 CGCGTTGGAATAACGGGCATTAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTGA 1434

QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCGAGGCTCGATATCAAAATTTGATCAGCTTTTAC 1650
DB 1435 AGCGACCTTTCAAGCAGGAGGATTAAGCAGAGAACGGCATTAAGCTTCTGCTCAGCTAC 1494

QY 1651 GCGCTCAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGGAGCATGACATGTTTGA 1710
DB 1495 GCCTTCGCGACATAGGCTCTGTTGACGACGCTCGAGGTCATCGGGCCAAATCCGTTAG 1554

QY 1711 GACACAGAAAGCTCTGGGACCGGTCTCTTTTAAATACGAGCAAAAGCTGGAGACACCG 1770
DB 1555 AAC---GAGAATTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAAGCTGAAGGCGCGC 1611

QY 1771 CTTTGTACTGATGGGACGGGATGACCGGATCGCGATCGGAGGCGGAGGAGGAGCTG 1830
DB 1612 ATACTCCTTAATCCACTCGCTTGGAGGACTTACCGCTGTCCGCTCGCAGGAGGAGCTTT 1671

; SEQUENCE 1, Application US/09609570
; Patent No. 6465204
; GENERAL INFORMATION:
; APPLICANT: Murphy et al.
; TITLE OF INVENTION: Amidases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CARELLA, BYRNE, BAIN, GILFILLAN,
; STREET: 6 BECKER FARM ROAD
; CITY: ROSELAND
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5 INCH DISKETTE
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WORD PERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/609,570
; FILING DATE: 30-Jun-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/427,372
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Charles J. Herron
; REGISTRATION NUMBER: 28,019
; REFERENCE/DOCKET NUMBER: 331400-53
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1869 NUCLEOTIDES
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: LINEAR
; MOLECULE TYPE: DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-609-570-1

Query Match 7.6%; Score 149.8; DB 4; Length 1869;
Best Local Similarity 52.3%; Pred. No. 3.1e-40;
Matches 356; Conservative 0; Mismatches 322; Indels 3; Gaps 1;

QY 1291 CCACCTATTCTTAACATACACGGCGGTCGGCATATGATGACGACATACATATTTTCAT 1350
DB 1135 CCGGTGATAGTCTTCGTCACGGCGGGCGGCGGAGGATGACGACACCGCTTCGTCTAC 1194

QY 1351 GAGTTTCAGTGTGGCGGGAAGGATACGGGTCTGTTTATATCAATCCGAGAGGAGC 1410
DB 1195 GAGATCAGGTGATGGCGGAGCAAGGCTACTACTGCTCTGCTGTAACCCGCGGCGCAG 1254

QY 1411 CACGGCTACGGGAGGATTTGTAATCGGTTCAGAGGAGATTATGGGGAAGGATTAT 1470
DB 1255 GACGGCTATACGGAAGACTTCGCGCTCGCGTCTCGGAGAGGACTGGCTTGGAGGACTTT 1314

QY 1471 GACGATGTGATGACGGTGTGGATGAGGTATCAACGAGATCGGCATATTTGATCCTAAG 1530
DB 1315 GAGGACATATGACGGCATCGAGGAGTCTTCAAGCTCGAACCGCGGCGGAGGAG 1374

QY 1531 CGGTCGGTGTACGGGCGGAGTACGGAGGTTTATGACCAACTGGATCGTGGGCGAG 1590
DB 1375 CGCGTTGGAATAACGGGCATTAAGCTACGGCGGCTTCATGACCAACTGGGCGCTTGA 1434

QY 1591 ACGAACCGCTTTAAAGCTGCGTTACCGAGGCTCGATATCAAAATTTGATCAGCTTTTAC 1650
DB 1435 AGCGACCTTTCAAGCAGGAGGATTAAGCAGAGAACGGCATTAAGCTTCTGCTCAGCTAC 1494

QY 1651 GCGCTCAGTATATCGGCTATTCTTTACAGACTGGCAGCTTGGAGCATGACATGTTTGA 1710
DB 1495 GCCTTCGCGACATAGGCTCTGTTGACGACGCTCGAGGTCATCGGGCCAAATCCGTTAG 1554

QY 1711 GACACAGAAAGCTCTGGGACCGGTCTCTTTTAAATACGAGCAAAAGCTGGAGACACCG 1770
DB 1555 AAC---GAGAATTCAGGAAGCTCAGCCCGCTGTCTACGCTCAGAAGCTGAAGGCGCGC 1611

QY 1771 CTTTGTACTGATGGGACGGGATGACCGGATCGCGATCGGAGGCGGAGGAGGAGCTG 1830
DB 1612 ATACTCCTTAATCCACTCGCTTGGAGGACTTACCGCTGTCCGCTCGCAGGAGGAGCTTT 1671
```



```
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 1010:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3085
US-09-221-017B-1010

Query Match          2.0%; Score 39.4; DB 4; Length 3085;
Best Local Similarity 51.6%; Pred. No. 0.016;
Matches 116; Conservative 0; Mismatches 106; Indels 3; Gaps 1;

QY 1347 TCATGAGTTTCAGGTGCTGGCGGCGGAAAGGATAGCGGTCGCTTTATATCAATCCGAGAGG 1406
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1716 TCCGAGACCCAGCTCTTCGCGAGCGCGGATATGCCACCTGCAAGTCAATTTCCGCGAT 1657
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1407 AAGCCACGGCTACGGCGAGGAATTTGTGAATCGGTGAGAGGATTTATGGGGGAAAGGA 1466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1656 CTCAGGCGGATACGCGAAGGAATTCCTCCGTGCGGATTCACACAGATCGGTGCGAAGC 1597
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1467 TTATGACATGTATGCGAGGCTGTGGATGAGGCTATCAACAGAGATCCGCATATTTGATCC 1526
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1596 CATGGACGATGGAGGACGGTGTGCGCTATGCTATCAGCCAAAGTTGGG---TGGATCC 1540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1527 TAAGCGGCTCGGTGTCACGGGCGGGAAGCTACGGAGTTTATGAC 1571
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1539 TGACAGGATCGCCATATACGGTGCCAGCCGCGGTGTTATGCCAC 1495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-08-834-776A-1
; Sequence 1, Application US/08834776A
; Patent No. 6060241
; GENERAL INFORMATION:
; APPLICANT: Cortesey-Theulaz, Irene
; TITLE OF INVENTION: Compositions and Methods Relating to
; TITLE OF INVENTION: Drug Discovery and Detection and Treatment of
; TITLE OF INVENTION: Gastrointestinal Diseases
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooley Godward Castro Huddleson & Tatum
; STREET: Five Palo Alto Square, 3000 El Camino Real
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94306-2155
; COMPUTER READABLE FORM:
```

```
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/834,776A
; FILING DATE: 03-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ashton, Nina M.
; REGISTRATION NUMBER: 37,273
; REFERENCE/DOCKET NUMBER: GAST-001/01US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-843-5000
; TELEFAX: 415-857-0663
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1395 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-834-776A-1

Query Match          1.8%; Score 35.8; DB 3; Length 1395;
Best Local Similarity 46.9%; Pred. No. 0.15;
Matches 112; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 898 TTGATCGGTGCTGTAACAGCGCCCGATTTTGGACAAAGGACAGCCAGGTTTATGTC 957
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 136 TTAGTGGCGGTTTTTGGGCTGTGCGGATACCCGAATACGCCATTGATTACATTATAAG 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 958 ATCGGCACAGATCAAGGCAGTACGGGCATCTATTATATTTTCATTGAAGGCCCTTGAT 1017
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 196 AAAGCATTAAGGATTTGATTCGTGAGCAATAATTGTGGCGTTGATTTGGGCTT 255
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1018 CGATTTCGTGGAAGAGAGTACATCAATAGCTTTCTCTTTCACCTCATCAACAGCAC 1077
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256 GCAATCTTTAGAAAAGAGAGATCAAAAAGATATCGCTTCGTATGCGGAGAAAT 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1078 TTTATGCGAGTGTGACAAAGCCGAGCAGCGAGTTCAGTTCAGTATCCGCTTGG 1136
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 316 AAGATTTTGAATCGCAATGCTGAACGGAGAAATTGAAGTCGTTTTCACCGCAAG 374
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihc
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 09:28:15 ; Search time 355 Seconds
(without alignments)
8666.800 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 1971

Sequence: 1 atgaaaagctgataacgcg.....catggttgatcaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1085931 seqs, 780495707 residues

Total number of hits satisfying chosen parameters: 2171862

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBseq.*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUBseq.*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUBseq.*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMBseq.*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUBseq.*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMBseq.*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUBseq.*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMBseq.*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUBseq.*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMBseq.*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUBseq.*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMBseq.*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUBseq.*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMBseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	807.8	41.0	1958	10	US-09-974-300-224
2	192.2	9.8	477	10	US-09-974-300-4866
3	149.8	7.6	1869	10	US-09-966-803-1
4	82.4	4.2	585	10	US-09-974-300-4838
5	46.4	2.4	1797	10	US-09-974-300-338
6	37	1.9	567	9	US-10-123-155-254
7	36.8	1.9	400	9	US-10-184-644-530
8	36.8	1.9	400	9	US-10-184-634-530
9	36.4	1.8	342	10	US-09-960-352-4863
10	36.4	1.8	1839	9	US-10-156-761-4548
11	36.4	1.8	9025608	9	US-10-156-761-1
12	36	1.8	568	10	US-09-974-300-2956
13	35.4	1.8	380	10	US-09-983-965-3736
14	35.4	1.8	3316	9	US-10-242-056-58
15	34.8	1.8	2374	9	US-10-237-271-6
16	34.8	1.8	2383	12	US-10-044-090-521
17	34.6	1.8	1141	9	US-10-184-644-120
18	34.6	1.8	1141	9	US-10-184-634-120
19	34.2	1.7	15150	9	US-10-146-473-9

20	34.2	1.7	15164	9	US-10-097-534-19	Sequence 19, Appl
21	34.2	1.7	15164	10	US-09-919-497-20	Sequence 20, Appl
22	33.6	1.7	444	9	US-10-123-155-498	Sequence 498, App
c 23	33.6	1.7	541	9	US-10-184-644-152	Sequence 152, App
c 24	33.6	1.7	541	9	US-10-184-634-152	Sequence 152, App
c 25	33.6	1.7	5828	9	US-09-436-184-6	Sequence 6, Appl
c 26	33.6	1.7	5828	10	US-09-903-248-6	Sequence 6, Appl
c 27	33.6	1.7	5828	10	US-09-859-604-6	Sequence 6, Appl
c 28	33.6	1.7	5828	10	US-09-903-063-6	Sequence 6, Appl
c 29	33.6	1.7	5828	10	US-09-903-216-6	Sequence 6, Appl
c 30	33.6	1.7	5828	10	US-09-903-199-6	Sequence 6, Appl
c 31	33.6	1.7	5828	10	US-09-880-107-3021	Sequence 3021, Ap
c 32	33.6	1.7	5828	10	US-09-903-023-6	Sequence 3, Appl
c 33	33.6	1.7	20300	9	US-09-939-209A-3	Sequence 6, Appl
34	33.2	1.7	837	9	US-10-184-644-352	Sequence 352, App
35	33.2	1.7	837	9	US-10-184-634-352	Sequence 352, App
36	33	1.7	567	9	US-10-123-155-476	Sequence 476, App
37	33	1.7	671	9	US-10-184-644-346	Sequence 346, App
38	33	1.7	671	9	US-10-184-634-346	Sequence 346, App
39	33	1.7	4108	10	US-09-925-301-493	Sequence 493, App
40	32.8	1.7	888	10	US-09-770-445-502	Sequence 502, App
41	32.4	1.6	504	9	US-09-918-995-9290	Sequence 9290, Ap
42	32.4	1.6	7680	10	US-09-939-581A-3	Sequence 3, Appl
43	32.2	1.6	514	9	US-09-918-995-32645	Sequence 32645, A
44	32.2	1.6	2706	10	US-09-815-242-5937	Sequence 5937, A
c 45	32	1.6	438	9	US-09-931-457A-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1

US-09-974-300-224
; Sequence 224, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 224
; LENGTH: 1958
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-224

Query Match	41.0%;	Score 807.8;	DB 10;	Length 1958;
Best Local Similarity	64.3%;	Pred. No. 6.5e-258;		
Matches 1268;	Conservative 0;	Mismatches 687;	Indels 18;	Gaps 3;
QY	1	ATGAAAAAGCTGATAACCGCAGACATCATCAGCATGTCTCTGTGACCGATGCTCAA	60	
DB	1	ATGAAGCAGCTGATAACCGAAAAAGACCTCATCAAGCTTGATTCATTACCATCCCGAG	60	
QY	61	TACGCCCGCAGCGGTACCCGTGCCCATATGTAATACAAAGTAAATCAAGAGAAAT	120	
DB	61	TATTCCCGCGATGTGTGAAAAATCGCTACGTCCAAACAAAGTAAAGAAAAACAGGAC	120	
QY	121	TCGTATACATCAATATATGATCTATGAAACGAAACGGGAGGATCTGCTCTCTTGGACA	180	
DB	121	TCATATGATTCGATATCATGATCTACGACCGGAAAAACAAAGCTTCGGTCAATGGAGC	180	
QY	181	CATGAGAAAAAGCAAGCAGCCAGCCCAAGATGGTCTCCGACGGCGACGCTTGCCTTT	240	

[illegible]

Qy	1318	CCGATATGATGTACGGACATACATATTTTCATGAGTTCTCAGTGCTGCGCGCAAGGA	1377
Db	1306	CCATACCCCATGTATCGAATGTATTTTCATGAATTTTCATGCTTGGCCGAAGGA	1365
Qy	1378	TACGCGTCTGTTATATCAATCCGAGAGAACCCACGGCTACGGCAGGAATTTGTGAAT	1437
Db	1366	AGCCCGCTGTFATATGTTAAACCCGCGGGGAAGCCACGGCTACGCACAGGATTTGTCAAC	1425
Qy	1438	CGGTCAGAGGAGNTATGGGGNAAGGATATACAGATGTGATGACGGCTGTGGATGAG	1497
Db	1426	CGGTTAGGGGGCACTACGGCGGTGGAGATTTTAAGGATGTATAGCGGCTGTTGACCAC	1485
Qy	1498	GCTATCAAAAGGATCCGCATATTGATTCCTAAGCGGCTCGGTCAGCGGCGGAAGCTAC	1557
Db	1486	GTGCTTGAACATTATGATTTTCGTCGATCAAGAAGGCTCGGCATCACCGGAGGAAGCTAC	1545
Qy	1558	GGAGTTTTATGACCACTGGATCGTCGGGCAGACGAACCGCTTAAAGCTGCGGTTACC	1617
Db	1546	GGGGGCTTTATGACAAACTGGCTGTGCGCCACACGAAGCGCTTCAAGCGCGCATGACA	1605
Qy	1618	CAGCGCTCGATATCAAAATTGGATCAGCTTTTCAGGGCGTCAGTGATATCGGCTATTCTTTT	1677
Db	1606	CAGAGTCCATTTTCCAACCTGGATCAGTGTTTTACGGGTAAGCGACATCTGGCTATTTTTTC	1665
Qy	1678	ACAGCTGGCAGCTTGAGCATGACATGTTTGGAGACACAGAAAAGCTCTGGGACCGGTCT	1737
Db	1666	ACAGCTGGAGCTCGGTGCAGACCTTTTGAAGATCCAGSCAACTGTGGAGCATTCG	1725
Qy	1738	CCTTTAAATACCGCAGCAACGTTGGAGACACCGCTTTTGATAGTCATGCGAGCGGAT	1797
Db	1726	CCGCTCAAAATACGGGCAACAGGTGGAGACCCCTCTCTCATCTTCGATGCGCAACGGGAC	1785
Qy	1798	GACCGATGCCCGATCGAGCGCGGACGCTGTTTATCGCTCTGAAAAAATGGCGAAG	1857
Db	1786	GACAGGTGCCGATTGAACAGGCGGAACAGCTGTTTACCGGCGCTGAAAAAATGGGAAA	1845
Qy	1858	GAACCAAGCTTGTCCGTTTCCGAATGCAATTCACCGACCGGACACCCA	1917
Db	1846	GAAGTAAACCTCGTCAGATTTCCGACGCGTCACACGATTTATCAAGAGCGGCATCCG	1905
Qy	1918	AGACAGCGGATCAAGCGCCTGAAATTATATCAGCTCATGTTTGTATCAACATCT	1970
Db	1906	AAGCAGCGGATCAGCGGCTTGAGTATTTGCAGGCTGTTGAGCTTATCT	1958

```

RESULT 2
US-09-974-300-4866
; Sequence 4866, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE OF INVENTION: Expression
; FILE REFERENCE: 10085:500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4866
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Bacillus clausii
US-09-974-300-4866

```

Query Match 9.8%; Score 192.2; DB 10; Length 477;
Best Local Similarity 65.1%; Pred. No. 4.3e-53;
Matches 300; Conservative 0; Mismatches 158; Indels 3; Gaps 1;

; PRIOR APPLICATION NUMBER: 09/580,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4838
; LENGTH: 585
; TYPE: DNA
; ORGANISM: Bacillus clausii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(585)
; OTHER INFORMATION: n = A,T,C or G
US-09-974-300-4838

Query Match 4.2%; Score 82.4; DB 10; Length 585;
Best Local Similarity 48.0%; Pred. No. 2e-16;
Matches 233; Conservative 0; Mismatches 252; Indels 0; Gaps 0;
QY 587 TTTCTCCTGACGGCAATGCTTGTCTTCTCAGCTAATTTAACTGAACAGATGATGCCA 646
Db 1 TTTTACGAATGACGCTGCTGTTCTTTGTAGCCGACGCCCAATAGGAAACCAATGATTTA 60
QY 647 GCAAGCGCATGATGTTTACATAATGTCACCTGGAGTCTGGAGATCTTAAAGCAGGTACAC 706
Db 61 CTTTTCGAGTTGATGTTTGTGTTTATGCTTGGATGCGGAGCAAAACGCAATACAA 120
QY 707 CTCATCGCGCTCATTCGGATCAAGCTCATTTTCCAGACGGAGGATGATCTTCTTTCG 766
Db 121 AGAACCAGTGGTTCGTAACAAAGCGGATCTGTATCACCGGATGTTCTTATGCTGCTCC 180
QY 767 TTGGAATGAAAGGATATAGAAATGCTAGCTCTCAAGGCGCTGCTCTATGATATCG 826
Db 181 TAGGCCATGAGAGAAATACAAACGCGACATGACGAAGCTGTATGTTTATGATGCC 240
QY 827 AACAAAGCCGCTCACATGCTTACTGAGATGCTGGAGCTTCATTTAGCGGATGCGCTGA 886
Db 241 AAAACAAGAAATCCGTTGCTTCCGCGCAAGTGGATGCTCAATTTGGCGATGTTGCAA 300
QY 887 TTGAGATTCATGATCGGTGCTCTACAGAGCGCCGATTTGGACAAGACAGCCCAAG 946
Db 301 TTGGGGAATTTTACAGAAATATGTTTCAATGGGTTGAAATGGGAGGAGATGTCAGT 360
QY 947 GGTATTATGTCATCGGCACAGATCAAGGACGATAGCGGATCTATATATTTTCGATGAAG 1006
Db 361 CCATTATGCGCTGTTTCAAAAGAGCGGATGTAATGTTTGGTTCGTTAGCCCTCGATG 420
QY 1007 GCCTTGTATPCGATGCTCTGGAAGAGAGTACATCAATAGCTTTTCTCTTCCACCTG 1066
Db 421 GCACAGTCAATCAGGTACGATGAAACAAACGTCACATAAATGGATTCGATTTTCAAAATG 480
QY 1067 ATGAA 1071
Db 481 GCCAA 485

RESULT 5
US-09-974-300-338
; Sequence 338, Application US/09974300
; Patent No. US20020146721A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods for Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27

; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 338
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-338

Query Match 2.4%; Score 46.4; DB 10; Length 1797;
Best Local Similarity 44.6%; Pred. No. 0.00042;
Matches 363; Conservative 0; Mismatches 436; Indels 15; Gaps 4;
QY 1094 CAAAGCCGACAGACCGAGTGTAGCTTTACAGTATCCCGCTTGGACAG---GAAGAGAAAC 1150
Db 902 CAAGCCGACAGCGCGACTATCTCTAGTATTAACGAAAGACAGTCAAAATGGCGG 961
QY 1151 AGCTGACTGGCGCGAATGACAAGTTTGTCTAGGAGCATACGATATCAATACCTGCAAGAGA 1210
Db 962 AATTAACGAAACATTCGGTTCCGGGTTCTCCGACACAGGAGCTGTCTTATCTCCGAGTTG 1021
QY 1211 TTCAATATGCTACAGAAGACGGGTGATGTTGAACGGCTGGCTGTAGTGGCCCTGCACAAA 1270
Db 1022 TCAGATATCTTCTTTTGTATGGCTCCCAATAGAGGAGTGTCTTTAAGCCCTCCCGG 1081
QY 1271 TGAAGGTGAGACAAATATCCACTTATTTCTTAACATACACGGGGTCCGCATATGATG 1330
Db 1082 TAGAAG-----CGAACGGCTGGACGATTTATTTGGCGCGCAGCGGACCGCAAGATCGAG 1135
QY 1331 ACGGACATACATATTTTCATGATGTTTCAGTGTGCGCGGCAAGAGTACGCGTCTGTTT 1390
Db 1136 AAACCTTATGTTCTACGATCTCTTCACTCCTGCAAAATGCGTTTACCAGTTATCG 1195
QY 1391 ATATCAATCCGAGAGAACCCAGCGCTACGGGCGAGAAATTTGTAATTCGGTTCAGAGGAG 1450
Db 1196 CTCCTAATTTTCAGGGGATCAGCGAATTCAGGTTATTCATTTCTTTAAATGGTTGAACAAG 1255
QY 1451 ATTATGGGGAAGGATTTATGAGATGTGATGAGCTGTGGATGAGGCTATCAACAGAG 1510
Db 1256 ATTGGGAGACGGCGCGCGCTCGATATGACCGGCGCATCGACTGCTTATCGATCAAA 1315
QY 1511 ATCCGCAATATGATCTTAAGCGGCTCGGTGTACGGCGGAAGACTACGGAGTTTATGA 1570
Db 1316 A---GCTTCAGACCGCGGAAACCTCTTTTAAATGGCGCGCATATGGAGGATATATG 1372
QY 1571 CCAACTGATGTCGGGCGAGACGACCGCTTTAAAGCTGCGTTTACCAGCGCTCGATAT 1630
Db 1373 CGCTTTTGTCTCCAGGGAGGCATCTCTAGATTTTCCGCGCGCTCGTATATATCGGAG 1432
QY 1631 CAAATTTGGATCAGCTTTACGGGCTCAGTATGATATCGGCTATTTCTTTACAGACTGGCAGC 1690
Db 1433 TCAGCAACCTGTTTTCGTTCTCAAAACCTACCTGATTTCTGGCAGCGGATGATGGAAA 1492
QY 1691 TTGAGCATGACATGTTTGA---GGACACAGAAAGCTCTGGGACCGGTCTCTCTTTAAAT 1747
Db 1493 AATGGTTCGCAATCCGAGCGGATTTATGAAAAAATGAAAGCGGATTTCTCTGTTACAT 1552
QY 1748 ACGCAGAAACGTGGAGACACCGCTTTTGATCTACTGCGCGCGGATGATCGCGATGCC 1807
Db 1553 ATCTGGAGAAATATGACTACGCGCATGCTGATCATCCAGGGCGCAACGATCCGCGCTCG 1612
QY 1808 CGATCGACAGCGGAGCAGCTGTTTATCGCTCTGAAAAAATGGGCAAGAAACCAAGC 1867
Db 1613 TAAAGGAGAGTCCGATCAGTCTCGATCAGTTGAGGAAACATGAGCGGGAATATCGAGT 1672
QY 1868 TTCTCCGTTTTCGAAATGCAATGCGACAAATTTATC 1901
Db 1673 ATCTCGTTTGGAAATGAAGGCGACGCGTTTC 1706

RESULT 6
US-10-123-155-254
; Sequence 254, Application US/10123155
; Publication No. US20030068794A1

APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HAITORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
PRIOR FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 1
LENGTH: 9025608
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: misc feature
LOCATION: (4187715)
OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 1.8%; Score 36.4; DB 9; Length 9025608;
Best Local Similarity 43.2%; Pred. No. 1.3e+02;
Matches 227; Conservative 0; Mismatches 296; Indels 3; Gaps 1;
QY 1304 ACATACACGGCGTCCGATATGATGATGACGACATACATATTTTCATGATGTTTCAGGTGC 1363
DB 5573387 ACATCACGGCGTCCGACCTGACGACGACGACTCGTTCGCGGGGCGCGGCGCT 5573328
QY 1364 TGGCGGGAAGGATACGCGGTTCGTTTATATCAATCCGAGAGAACCCAGCGTACGGGC 1423
DB 5573327 GCGTGACACCGGTACGCGGTTCGTCGCGGTCACTACCGCGGTGACCGGATACGGGC 5573268
QY 1424 AGGAATTTGTGATGCGGTACAGAGGATTTATGGGGAAAGGATTTATGACGATGTGATGC 1483
DB 5573267 GGGAGTGGACGACGCGCTCAAGCACCGGTGGTCTGATCGAGCTGGAGGACATCGCGG 5573208
QY 1484 AGGCTGTGATGAGGTATCAACAGGATCCGATATTCATCTTAAGCGGCTCGGTGTC 1543
DB 5573207 CGGTGCGGAGTGGCGGTGTCGTCGCGGTCCGCGACCCGACAGGCTGATCCTGACCG 5573148
QY 1544 CGGGCGGAAGCTACGAGGTTTTATGACCAACTGGATCGTCGGGACAGACGAAACCGCTTTA 1603
DB 5573147 GCGGTTCTCGGGCGGTATCTCACGCTCGGCTCGGACCCAGCGGAGCGGTGA 5573088
QY 1604 AGCTGCGGTTACCGAGCGCTCGATATCAAAATGGATCAGCTTTCACGCGCTCAGTGATA 1663
DB 5573087 CGATCGGGATCGCGGTCGCGGTGGCGGACTACGTCACGCGGTACCCAGCAGATGG 5573028
QY 1664 TCGGCTATTTTACAGCTGGCAGCTTGACATGACATGTTTGAAGNCA---CAGAAA 1720
DB 5573027 AGCCCTGAAGCGGTGACCGGACGCTGCTCGGGGCGACCGCGGAGGTGCGCGAGC 5572968
QY 1721 AGCTCGGACCGGTCTCTTTTAAATACGACGAAACGCTGGAGACACCGCTTTTGATAC 1780
DB 5572967 GGTTCGAGCGGTGCGCGGTGACCTACGTGACGCGGTGAAGCACCAGGTGTACATCT 5572908
QY 1781 TGCATGGCGGATGACCGATCCCGATGACGACGCGGAGCA 1826
DB 5572907 CGGCGGCGGTCAACGATCCGCGGTGCCCATCCGCGGTGGAGAA 5572862

RESULT 12
US-09-974-300-2956
Sequence 2956, Application US/09974300
Patent No. US20020146721A1
GENERAL INFORMATION:
APPLICANT: Berka, Randy M.
APPLICANT: Clausen, Ib Groth

TITLE OF INVENTION: Methods For Monitoring Multiple Gene
TITLE OF INVENTION: Expression
FILE REFERENCE: 10085.500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2956
LENGTH: 568
TYPE: DNA
ORGANISM: Bacillus licheniformis
US-09-974-300-2956

Query Match 1.8%; Score 36; DB 10; Length 568;
Best Local Similarity 47.4%; Pred. No. 0.57; Indels 0; Gaps 0;
Matches 108; Conservative 0; Mismatches 120;
QY 392 AGGGGAAAGCATTTGATGACCGAGAAAAACAGACGACGATGATGAACCTGTTGAAG 451
DB 317 ATGAAGAAAGGCTCGTCTTCATCAGAAAGCTGTTGGATTTTGGACCATCCGTTT 376
QY 452 TGCAAGCCCTCTCTACAAACGGGACCGCAAGGGCTGACGAGAGGTGCGGTATGCCAGC 511
DB 377 TGAAGGACGACGACATCATCTGACGATCATCTGACGAGCGGACGCGCCGACGCC 436
QY 512 TTGCTGTGTGTCAGCGTAAAGTCGGGTGAGATGAAGAGCTGACAAGTCACAAAGCTGATC 571
DB 437 TCCTCATCTTCTCGATTTCCACGGCGTATTCACAAAGCTGACCGGTATCAAGACTTTC 496
QY 572 ATGCTGATCTGCTTTTCTCTGACGGCAAAATGGCTGTTTCTCTAG 619
DB 497 ATCACTACTGGCGGACGTTTATCAAAATCAAGCGGTTGATGTAACAG 544

RESULT 13
US-09-983-965-3756/c
Sequence 3756, Application US/09983965
Patent No. US20020137160A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION
FILE REFERENCE: 37-21(10297)C
CURRENT APPLICATION NUMBER: US/09/983,965
CURRENT FILING DATE: 2001-10-26
PRIOR APPLICATION NUMBER: US 09/465,231
PRIOR FILING DATE: 1999-12-15
PRIOR APPLICATION NUMBER: US 60/113,678
PRIOR FILING DATE: 1998-12-17
NUMBER OF SEQ ID NOS: 5912
SEQ ID NO 3756
LENGTH: 380
TYPE: DNA
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: unsure
LOCATION: (371)
OTHER INFORMATION:
OTHER INFORMATION: Clone ID: 46-LIB3058-008-Q1-K1-D10
US-09-983-965-3756

Query Match 1.8%; Score 35.4; DB 10; Length 380;
Best Local Similarity 56.4%; Pred. No. 0.69; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 51;
QY 87 ATATGAAATCAAGTAATCAAGAGAAAGATTGTTATACATCAATATATGATCTA 146

Db 130 AAAATTAACAAACCTAATGATCTTAAGAGATATGAGCATAAATCCGAGAGATGAATCTG 71
QY 147 TGAACGAAACGAGGAGGATCTGTTCTTGACACATGAGAGAAAGCGAACGACCGA 203
Db 70 GGAAGCAAAACTGTGTCATCTGCTCGATTATATAAGAAATAAATCCCGGA 14

RESULT 14

US-10-242-056-58
; Sequence 58, Application US/10242056
; Publication No. US20030113323A1
; GENERAL INFORMATION:
; APPLICANT: Ensign, Jerald C
; APPLICANT: Bowen, David J
; APPLICANT: Petell, James
; APPLICANT: Fatig, Raymond
; APPLICANT: Schoonover, Sue
; APPLICANT: ffrench-Constant, Richard
; APPLICANT: Orr, Gregory L
; APPLICANT: Merlo, Donald J
; APPLICANT: Roberts, Jean L
; APPLICANT: Rocheleau, Thomas A
; TITLE OF INVENTION: Insecticidal Protein Toxins from
; TITLE OF INVENTION: Photorhabdus
; NUMBER OF SEQUENCES: 88

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dowelanco
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: IN
; COUNTRY: US
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/242,056
; FILING DATE:

; CLASSIFICATION:
; APPLICATION NUMBER: US 08/063,615
; FILING DATE: 18-MAY-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/395,497
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/007,255
; FILING DATE: 06-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/608,423
; FILING DATE: 28-FEB-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/705,484
; FILING DATE: 28-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/743,699
; FILING DATE: 06-NOV-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Borucki, Andrea T.
; REGISTRATION NUMBER: 33651
; REFERENCE/DOCKET NUMBER: 50301E
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-337-4846
; TELEFAX: 317-337-4847
; INFORMATION FOR SEQ ID NO: 58:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4932 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)
US-10-242-056-58
Query Match 1.8%; Score 35.4; DB 9; Length 3316;
Best Local Similarity 13.8%; Pred. No. 2.9;
Matches 46; Conservative 125; Mismatches 162; Indels 0; Gaps 0;
QY 1272 GGAAGGTGACACATATCCACTTATTTTAAACATACACGCGGTCCGCATATGATGA 1331
Db 2164 RGASNGU(GYTRAAAAR)UUAATHRUHSHRVATHRVAVSGYSRTYRAA)TRGUGUGUT 2223
QY 1332 CGGACATACATATTTTCATGAGTCTTCAGTCTGCGCGGCGGAAAGATACGCGGTGTTTA 1391
Db 2224 HRTHRGYTVRASNUY(((RRASGY)THRVAUUASTRHASYSASNHAAGYU)ASNYUG 2283
QY 1392 TATCAATCCGAGAGAGAGCCAGCGGTACGGCAGCAATTTGTGAATCGCGTCAGAGGAGA 1451
Db 2284 USRVAHTRSRRASTRTRTHR)TH((RYSASNHSRYSASASNARGYSHTYRGN)G 2343
QY 1452 TTATGGGGAAAGGATTATGACGATGTGATGCGAGGCTGTGGATGAGGCTATCAACGAGA 1511
Db 2344 UASNAAGUTHRAAASGYARGASNUHYSARGTYR)SRT(HRCNTHRHGYUTHRSRGYAAATH 2403
QY 1512 TCGCATATTGATCTTAAGCGGCTCGGTGTCTACGCGGCGGAAAGCTACGAGGTTTATGAC 1571
Db 2404 RTYRSRTHR)TYRTHRUSRGUAAASHSRTHRASRYSASNTYR((UGNVACYSUAS 2463
QY 1572 CAACTGGATCGTGGCGACGAGCAACCGCTTTAA 1604
Db 2464 NVAVATRASHSTYNASARGSR)GYISYSGYAA 2496

RESULT 15

US-10-237-271-6
; Sequence 6, Application US/10237271
; Publication No. US20030096328A1
; GENERAL INFORMATION:
; APPLICANT: THE BURNHAM INSTITUTE
; APPLICANT: SMITH, Jeffrey W.
; APPLICANT: KRIDEL, Steven J.
; APPLICANT: AXELROD, Fumiko T.
; TITLE OF INVENTION: SERINE/THREONINE HYDROLASE PROTEINS AND SCREENING ASSAYS
; FILE REFERENCE: BURN100-1
; CURRENT APPLICATION NUMBER: US/10/237,271
; CURRENT FILING DATE: 2002-09-04
; PRIOR APPLICATION NUMBER: US 60/317,842
; PRIOR FILING DATE: 2001-09-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 2374
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-237-271-6

Query Match 1.8%; Score 34.8; DB 9; Length 2374;
Best Local Similarity 49.5%; Pred. No. 3.7;
Matches 90; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 1721 AGCTCTGGGACCGGTCTCCTTTTAAATACGCAACAAACGTGGAGACACCGCTTTTGATAC 1780
Db 1968 AGATGCTGGCAAAATCGCCCATCAGATACATCCCTCAGTGAAGACACCACTGTTACTGA 2027
QY 1781 TGCATGGCGAGCGGGATGACCGATGCCGATCGAGCGGAGGAGAGCTGTTATCGCTC 1840
Db 2028 TGTTGGGCGAGGAGGACCGCGGTGTGCCCTTCAAGCAGCGCATGGAGTATTACCGTGCC 2087
QY 1841 TGAAAAAATGGCAAGAAACCAACCAAGCTTGCTCGTTTCCGATGATCGACCAATTTAT 1900
Db 2088 TCAAGACCCCGAATGTGCTGTTCGGCTCCTCTCTATCCCAAGACCAACCCACCGCATAT 2147
QY 1901 CA 1902

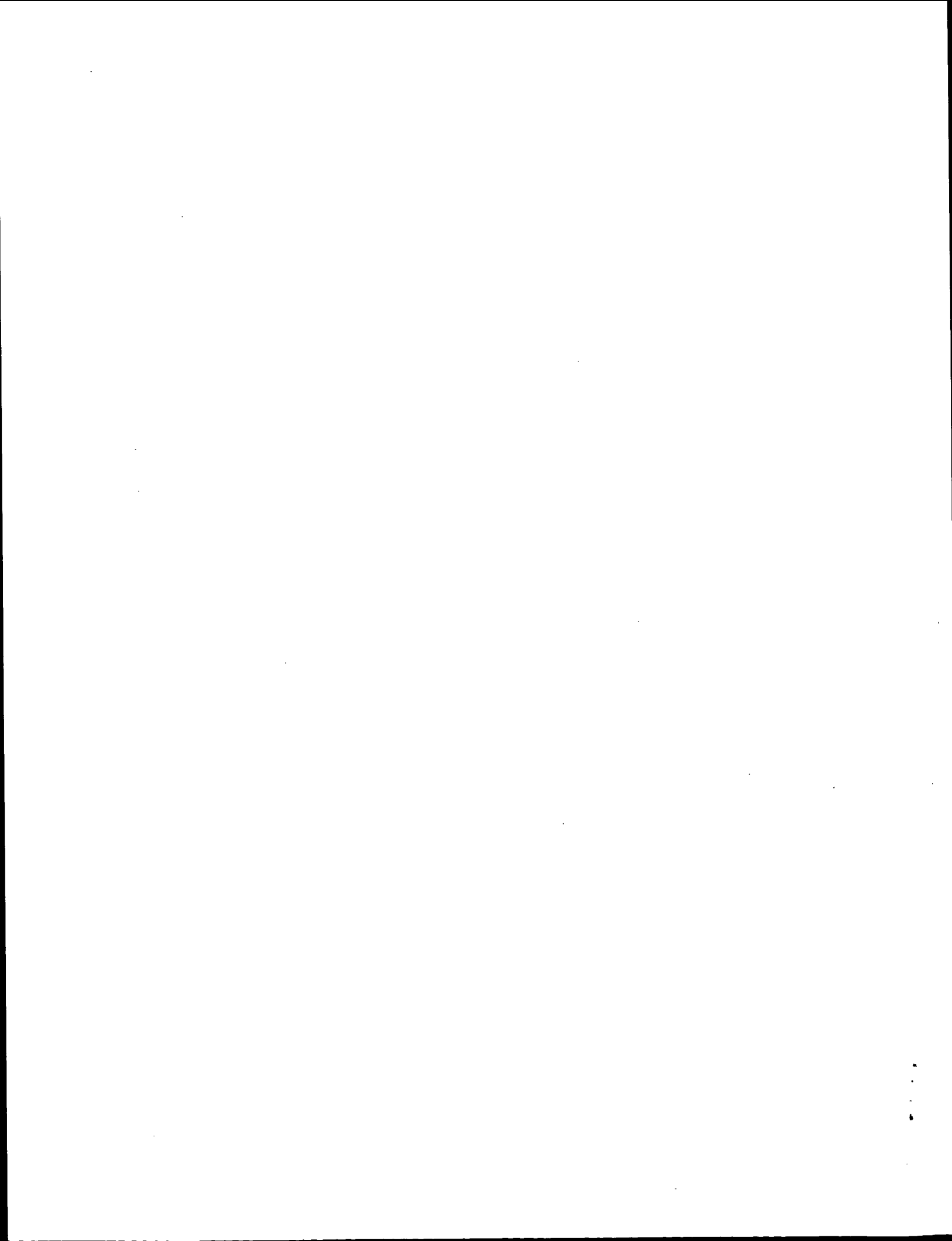
Sat Jul 5 15:57:28 2003

us-09-462-845-1.rnpb

Page 9

Db 2148 CA 2149

Search completed: July 4, 2003, 11:48:54
Job time : 370 secs



GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 4, 2003, 11:05:37 ; Search time 2559 Seconds
(without alignments)
12474.132 Million cell updates/sec

Title: US-09-462-845-1

Perfect score: 171

Sequence: 1 atgaaaagctgataaccgc.....catggttgatcaacatctc 1971

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_estc: *
9: gb_estl: *
10: gb_est2: *
11: gb_hc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	66	3.3	839	17	AZ535741
C 2	66	3.3	909	17	AZ530780
C 3	50.4	2.6	552	13	BM175145
C 4	48.4	2.5	327	9	AA071581
C 5	46.8	2.4	475	12	BG279712
C 6	43.6	2.2	697	13	BJ153314

C 7	43	2.2	840	13	BJ118680
C 8	42.6	2.2	873	17	AZ528669
C 9	42.6	2.2	939	12	BG025910
C 10	42	2.1	695	13	BJ133826
C 11	40.8	2.1	490	13	BJ481303
C 12	40.8	2.1	603	10	AV836520
C 13	40.8	2.1	1101	17	CNS00001
C 14	39.8	2.0	531	12	BG650557
C 15	39.6	2.0	951	11	AY105542
C 16	39.6	2.0	1449	14	BQ049806
C 17	39.4	2.0	711	10	BE130860
C 18	39.4	2.0	1064	10	BE036593
C 19	37.6	1.9	719	14	BQ443031
C 20	37.4	1.9	1101	17	CNS003EJ
C 21	37.2	1.9	513	10	AW906913
C 22	37.2	1.9	884	17	AZ540694
C 23	37.2	1.9	926	17	AZ689494
C 24	37.2	1.9	932	17	AZ692448
C 25	37.2	1.9	956	17	CNS007BU
C 26	37	1.9	438	13	BM570134
C 27	37	1.9	459	9	AA792627
C 28	37	1.9	491	13	BJ169948
C 29	37	1.9	532	10	AV601962
C 30	37	1.9	579	13	BM565239
C 31	37	1.9	583	13	BM506840
C 32	37	1.9	583	13	BM565238
C 33	37	1.9	588	12	BG800737
C 34	37	1.9	725	13	BI555774
C 35	37	1.9	733	13	BI730952
C 36	37	1.9	747	13	BG971297
C 37	37	1.9	935	14	BQ920566
C 38	37	1.9	939	14	BQ957767
C 39	37	1.9	1000	12	BF534027
C 40	36.8	1.9	377	10	AW353260
C 41	36.8	1.9	406	10	AW353257
C 42	36.6	1.9	717	17	BH028231
C 43	36.4	1.8	418	10	AV590668
C 44	36.4	1.8	457	12	BF707325
C 45	36.4	1.8	526	10	AV605765

ALIGNMENTS

RESULT 1
AZ535741/c
LOCUS
DEFINITION
ACCSSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AZ535741
ENTCC88TR Entamoeba histolytica Sheared DNA
genomic, DNA sequence.
AZ535741
GI:11092688
Entamoeba histolytica.
Entamoeba histolytica.
Fukaryota; Entamoebidae; Entamoeba.
1 (bases 1 to 839)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
HM1:IMSS sheared DNA library
Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 34
High quality sequence stop: 716.

FEATURES
 source
 Location/Qualifiers
 1. 839
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."
 277 a 143 c 107 g 312 t
BASE COUNT
ORIGIN
 Query Match 3.3%; Score 66; DB 17; Length 839;
 Best Local Similarity 48.6%; Pred. NO. 5e-08;
 Matches 214; Conservative 0; Mismatches 220; Indels 6; Gaps 1;
 QY 1147 AACAGCTGACTGGCGGAATGACAAAGTTTGTTCAGGGAGCATACATATCAATACCTGAA 1206
 DB 569 AACAAATCACTACATTAATCAAGAAGTTCTTCTACAATTAATTTGGAGAATATAAA 510
 QY 1207 GAGATTCAATATGTACAGAGAGCGGCTGATGTGAACGGCTGCTGATGAGGCTGCA 1266
 DB 509 GAAATTCATTACACTGGAGCAATTAATGATCAAAATTCATGCAATTTACTTACCCACCA 450
 QY 1267 CAATGGAGGTGAGACAAATATCCACTTATTCATACATACACGGCGGCTCGCATATG 1326
 DB 449 AATATGAACAAACACAAATATCCAGTAATTTTATATACACATGGAGGACCAATCA 390
 QY 1327 ATGTACGGACATACATATTTTCAT-----GAGTTTTCAGGTGCTGGCGCGAAAGGATAC 1380
 DB 389 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAGTTTATAGCACACAAAGTTAT 330
 QY 1381 GCGTFCGTTTATATCAATCCGAGAGAGAGCCACGGCTAGCGGCGAGGAATTTGTAATGCG 1440
 DB 329 ATTGTTTGTGACCAAAATTTCCATGGATCAGGAAGTTATGAGATGCATTTTAAAGCA 270
 QY 1441 GTCAGAGGATTTATGGGGAAGGATATGACGATGTCATGAGGCTGTGGATGAGGCT 1500
 DB 269 ATTAGAAGAAATTTGGGAGGATGGCCATTGGAAGATTTAATGAAGGAATGGAATTTTA 210
 QY 1501 ATCAACAGAGATCCGCATATTTGATCCTTAAGCGGCTGGTGTACGCGGCGAAAGCTACGA 1560
 DB 209 AAACTAGTGAGCCATTAGTTGATATAGATAATGCTGCTGCTATGGGCAAGTTATGGA 150
 QY 1561 GGTTTATGACCAACTGGAT 1580
 DB 149 GGATATGATGATGATTTGGAT 130
RESULT 2
 AZ530780/c 909 bp DNA linear GSS 03-NOV-2000
 LOCUS ENTBV92TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
 DEFINITION
 AZ530780
 ACCESSION AZ530780.1 GI:11084914
 VERSION
 KEYWORDS Entamoeba histolytica.
 SOURCE Entamoeba histolytica
 ORGANISM
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCES
 1 (bases 1 to 909)
 Loftus B., Van Aken S. and Fraser C.
 Determination of clone end sequences from Entamoeba histolytica
 TITLE

JOURNAL
 COMMENT
 HMI:IMSS sheared DNA library
 Unpublished (2000)
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: bjoftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 15
 High quality sequence stop: 780.
FEATURES
 source
 Location/Qualifiers
 1. 909
 /organism="Entamoeba histolytica"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: PHOS1; Site.1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."
 299 a 151 c 118 g 341 t
BASE COUNT
ORIGIN
 Query Match 3.3%; Score 66; DB 17; Length 909;
 Best Local Similarity 48.6%; Pred. NO. 5.2e-08;
 Matches 214; Conservative 0; Mismatches 220; Indels 6; Gaps 1;
 QY 1147 AACAGCTGCTGGCGGAATGACAAAGTTTGTTCAGGGAGCATACGATATCAATACCTGAA 1206
 DB 665 AACAAATCACTACATTAATCAAGAAGTTCTTCTACAATTAATTTGGAGAATATAAA 606
 QY 1207 GAGATTCAATATGTACAGAGAGCGGCTGATGTGAACGGCTGCTGATGAGGCTGCA 1266
 DB 605 GAAATTCATTACACTGGAGCAATTAATGATCAAAATTCATGCAATTTACTTACCCACCA 546
 QY 1267 CAATGGAGGTGAGACAAATATCCACTTATTCATACATACACGGGCTGCGCATATG 1326
 DB 545 AATATGAACAAACACAAATATCCAGTAATTTTATATACATGGAGGACCAATCA 486
 QY 1327 ATGTACGGACATACATATTTTCAT-----GAGTTTTCAGGTGCTGGCGCGAAAGGATAC 1380
 DB 485 CCATGGACTAATAATTTCCATTATAGATGGAATCCACAAAGTTTATAGCACCAAGTTAT 426
 QY 1381 GCGTFCGTTTATATCAATCCGAGAGAGAGCCACGGCTAGCGGCGAGGAATTTGTAATGCG 1440
 DB 425 ATTGTTTGTGACCAAAATTTCCATGGATCAGGAAGTTTATGGAGATGCATTTTAAAGCA 366
 QY 1441 GTCAGAGGATTTATGGGGAAGGATTTATGACGATGTCATGAGGCTGTGGATGAGGCT 1500
 DB 365 ATTAGAAGAAATTTGGGAGGATGGCCATTGGAATTTAATGAAGGAATGGAATTTTA 306
 QY 1501 ATCAACAGAGATCCGCATATTTGATCCTTAAGCGGCTGGTGTACGCGGCGAAAGCTACGA 1560
 DB 305 AAACTAGTGAGCCATTAGTTGATATAGATAATGCTGCTGCTATGGGCAAGTTATGGA 246
 QY 1561 GGTTTATGACCAACTGGAT 1580
 DB 245 GGATATGATGATGATTTGGAT 226

Db 142 ATTAATGCGGCAAGGGTACTATATAGTTTACGTTAAATCCAGGGGAGCAATGGATA 83
 QY 1419 CGGGCAGGAATTTGTGTAATCGGTCAGAGGAGATATTGGGGAAAGGATTTATGAGATGT 1478
 Db 82 TAGTGAAGACTTTGCACATAGAGATATTAAACAGAACTGGACTGGAGGACTTCCAAGATAT 23
 QY 1479 GATCAGGCTGTGATCA 1496
 Db 22 ATTAACGGGTATTGAGGA 5

RESULT 5
 BG279712
 LOCUS
 DEFINITION
 b5e08np.r1 Neurospora crassa sexual cDNA library, Uni-zap vector
 system Neurospora crassa cDNA clone b5e08np 5', mRNA sequence.
 ACCESSION
 BG279712
 VERSION
 BG279712.1 GI:13077352
 KEYWORDS
 EST.
 SOURCE
 Neurospora crassa.
 ORGANISM
 Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariales; Sordariaceae; Neurospora.
 REFERENCE
 1 (bases 1 to 475)
 Kupper, D., Lai, H., Nelson, M. and Roe, B.
 ESTs from a Neurospora crassa Sexual cDNA Library
 Unpublished (2001)
 TITLE
 Other_ESTs: b5e08np.f1
 JOURNAL
 COMMENT
 Contact: Bruce A. Roe, University of Oklahoma, broe@ou.edu
 Department of Chemistry and Biochemistry
 Advanced Center for Genome Technology, University of Oklahoma
 620 Parrington Oval, Norman, OK 73019, USA
 Tel: 405 325 4912
 Fax: 405 325 7762
 Email: broe@ou.edu
 Contact Dr. Mary Anne Nelson, Department of Biology, University of
 New Mexico, Albuquerque, NM 87131 (e-mail address manelson@unm.edu)
 regarding clone availability
 Seq primer: M13 Universal Reverse Primer
 High quality sequence stop: 474.
 Location/Qualifiers
 1. 475
 /organism="Neurospora crassa"
 /strain="wild type"
 /db_xref="taxon:5141"
 /clone="b5e08np"
 /clone_lib="Neurospora crassa sexual cDNA library, Uni-zap
 vector system"
 /tissue_type="perithecia (fruiting bodies)"
 /dev_stage="sexual"
 /lab_host="E. coli strain SOLR"
 /note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
 XhoI; 5' end of cDNA cloned into EcoRI site of pBluescript
 ; 3' end of cDNA cloned into XhoI site of pBluescript"
 BASE COUNT 103 a 120 c 146 g 105 t 1 others
 ORIGIN

FEATURES

source
 Query Match 2.4%; Score 46.8; DB 12; Length 475;
 Best Local Similarity 52.8%; Pred. No. 0.02;
 Matches 122; Conservative 0; Mismatches 108; Indels 1; Gaps 1;
 QY 1360 GTGCTGGCGGGAAGGATACGCGTCTTTTATATCAATCCGAGAGAGCCACGCTAC 1419
 Db 151 GTCCTTGGGAGCAGGCTAGCTGCTGTGAGCCCGAACCCGACAGGAGTACGGTTAT 210
 QY 1420 GGGCAGGAATTTGTGAA -TGGCGTCAGAGGAGATTTATGGGGAAAGGATTTATGAGATCT 1478
 Db 211 GGCATGGCTCTCCAAAACGGNAATCAAGGGCCAGTGGGTGGCGGCGGTATGAGATCT 270
 QY 1479 GATCAGGCTGTGATGAGGCTATCAACAGGATCCGATATTGATCTTAAGCGCTCGG 1538
 Db 271 GGTCAAAGCGTTTGAGCATATTGAGGAGAACATGCCGTACGTAGATACGAGGCGGT 330

QY 1539 TGTACGGCGGAAGCTACGGAGGTTTATGACCAACTGGATCGTCGGGCA 1589
 Db 331 CCATTTGGGTGCTAGTTATGTTGGTTATATGATCAATTTGGATCAAGGCCA 381
 RESULT 6
 BJ153314/c
 LOCUS
 DEFINITION
 BJ153314 unpublished oligo-capped cDNA library, C. elegans L1 stage
 Caenorhabditis elegans cDNA clone yk1318a03 3', mRNA sequence.
 ACCESSION
 BJ153314
 VERSION
 BJ153314.1 GI:18321299
 KEYWORDS
 EST.
 SOURCE
 Caenorhabditis elegans.
 ORGANISM
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE
 1 (bases 1 to 697)
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
 and Sugano, S.
 A complementary view of the C. elegans genome
 Unpublished (2002)
 JOURNAL
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1. 697
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1318a03"
 /clone_lib="unpublished oligo-capped cDNA library, C.
 elegans L1 stage"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

BASE COUNT 212 a 151 c 162 g 171 t 1 others
 ORIGIN
 Query Match 2.2%; Score 43.6; DB 13; Length 697;
 Best Local Similarity 51.3%; Pred. No. 0.23;
 Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 QY 1375 GGATCGCGTCTGTTATATCAATCCGAGAGAGCCACGCTACGGGACGAAATTTGTG 1434
 Db 696 GGATATGCCCTCWTCAAGTCAATTCGCTGGATCACTCGGATTTGGTACGATTTATT 637
 QY 1435 AATGCGGTCAAGAGAGATTTATGGGGAAAGGATTTATGACGATGTGATCGAGCTGTGAT 1494
 Db 636 CGTGCTCTGCTGGAACACTCGGAGATATGGATGTTAAGGATGTTTCATAACGCGATCTC 577
 QY 1495 GAGGCTATCAACAGAGATCCGCATATTGATCCTTAAGCGGTCGGTGTACCGGCGGAAC 1554
 Db 576 ACCGTCTTGGATGAAGATCCAGCTATTTCTCGAGACAAGGTTTACTCTTCGGAGGATCT 517
 QY 1555 TACGGAGGTTTATG 1569
 Db 516 CATGGAGGTTCTCTG 502

RESULT 7
 BJ118680/c
 LOCUS
 DEFINITION
 BJ118680 unpublished oligo-capped cDNA library, C. elegans L1 stage
 Caenorhabditis elegans cDNA clone yk1318a03 3', mRNA sequence.
 ACCESSION
 BJ118680
 VERSION
 BJ118680.1 GI:18321299
 KEYWORDS
 EST.
 SOURCE
 Caenorhabditis elegans.
 ORGANISM
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE
 1 (bases 1 to 697)
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
 and Sugano, S.
 A complementary view of the C. elegans genome
 Unpublished (2002)
 JOURNAL
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1. 697
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1318a03"
 /clone_lib="unpublished oligo-capped cDNA library, C.
 elegans L1 stage"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

BASE COUNT 212 a 151 c 162 g 171 t 1 others
 ORIGIN
 Query Match 2.2%; Score 43.6; DB 13; Length 697;
 Best Local Similarity 51.3%; Pred. No. 0.23;
 Matches 100; Conservative 0; Mismatches 95; Indels 0; Gaps 0;
 QY 1375 GGATCGCGTCTGTTATATCAATCCGAGAGAGCCACGCTACGGGACGAAATTTGTG 1434
 Db 696 GGATATGCCCTCWTCAAGTCAATTCGCTGGATCACTCGGATTTGGTACGATTTATT 637
 QY 1435 AATGCGGTCAAGAGAGATTTATGGGGAAAGGATTTATGACGATGTGATCGAGCTGTGAT 1494
 Db 636 CGTGCTCTGCTGGAACACTCGGAGATATGGATGTTAAGGATGTTTCATAACGCGATCTC 577
 QY 1495 GAGGCTATCAACAGAGATCCGCATATTGATCCTTAAGCGGTCGGTGTACCGGCGGAAC 1554
 Db 576 ACCGTCTTGGATGAAGATCCAGCTATTTCTCGAGACAAGGTTTACTCTTCGGAGGATCT 517
 QY 1555 TACGGAGGTTTATG 1569
 Db 516 CATGGAGGTTCTCTG 502

RESULT 7
 BJ118680/c
 LOCUS
 DEFINITION
 BJ118680 unpublished oligo-capped cDNA library, C. elegans L1 stage
 Caenorhabditis elegans cDNA clone yk1318a03 3', mRNA sequence.
 ACCESSION
 BJ118680
 VERSION
 BJ118680.1 GI:18321299
 KEYWORDS
 EST.
 SOURCE
 Caenorhabditis elegans.
 ORGANISM
 Caenorhabditis elegans
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea
 ; Rhabditidae; Peloderinae; Caenorhabditis.
 REFERENCE
 1 (bases 1 to 697)
 Kohara, Y., Shin-i, T., Thierry-Mieg, J., Thierry-Mieg, D., Suzuki, Y.
 and Sugano, S.
 A complementary view of the C. elegans genome
 Unpublished (2002)
 JOURNAL
 COMMENT
 Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshin@genes.nig.ac.jp.
 Location/Qualifiers
 1. 697
 /organism="Caenorhabditis elegans"
 /strain="N2"
 /db_xref="taxon:6239"
 /clone="yk1318a03"
 /clone_lib="unpublished oligo-capped cDNA library, C.
 elegans L1 stage"
 /sex="hermaphrodite"
 /tissue_type="whole animal"
 /dev_stage="L1"
 /note="The AD-wrmcDNA library was generated with poly(A)+
 RNA isolated from both hermaphrodite and male N2 worms of
 all larval stages, embryos, adults and dauers and the
 subsequent generation of cDNAs by poly(A) priming. The
 cDNAs were cloned into pPC86"

source

```
1. .951
/organism="Zea mays"
/db_xref="MaizeDB:635221"
/db_xref="taxon:4577"
/clone="PC0091746"
/clone_lib="Maize Mapping Project/DuPont Cornsensus
Library"
/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed DuPont contigs; this resource was
assembled by DuPont as part of a collaboration for the
overgo addressing of BACs in conjunction with the Maize
Mapping Project"
BASE COUNT      259 a      207 c      207 g      278 t
ORIGIN
```

```
Query Match      2.08; Score 39.6; DB 11; Length 951;
Best Local Similarity 47.78; Pred. No. 4.4;
Matches 148; Conservative 0; Mismatches 159; Indels 3; Gaps 1;

QY 1283 CAACATATCCACTTATTCTTAACATACACGGCGTCCGCATATGATCTACGGACATACAT 1342
    ||| |||| || || || || || || || || || || || || || || || || || ||
Db 317 CAAGTAATCCAAACAATCGTTGTTCTTCATGGTGGACACACTCAGTTTATCCATCAAGTT 376

QY 1343 ATTTTCATGAGTTTCAGTGTGCGGCGGAAAGGATACGCGTCTGTTTATATCAATCCGA 1402
    ||| || || || || || || || || || || || || || || || || || || ||
Db 377 ATTCCAAATCCTTGGCTTTCTTTTTCGACGGGATATAATCTTCTTGTGTAATACTA 436

QY 1403 GAGGAAGCCACGCGTACGGCAGGAAATTTGTGAATCGGTCAGAGAGATTATGGGGAA 1462
    ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 437 GGGTTTCATTAGCTTTGGGGAGAGAGACACTGCAATCCCTTCTCGCAATATTGGTTCTC 496

QY 1463 AGGATTATGACGATGTGATGCGAGCTGTGGATGAGGCTATCAACGAGATCCGCATATTG 1522
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 497 AGGATGTCATATGATGATTGACGGCATTGGACCTTGTATTAAAGAGAGACTA---ATAG 553

QY 1523 ATCCTAAGCGGCTCGTGTACACGGCGGGAAGCTACGGAGTTTATGACCAACTGGATCG 1582
    |||| || || || || || || || || || || || || || || || || || || ||
Db 554 ATCCATCCAGAGTTCCCGTAGTTGGAGGTTTCACATGGAGGTTTCTCTGACACACATTTGA 613

QY 1583 TGGGCAGAC 1592
    || || || ||
Db 614 TTGCCAGGC 623
```

Search completed: July 4, 2003, 14:02:53
Job time : 2565 secs

